CHAPTER 4
The Test Characteristic Curve

CHAPTER 4

The Test Characteristic Curve

Item response theory is based upon the individual items of a test, and up to this point, the chapters have dealt with the items one at a time. Now, attention will be given to dealing with all the items in a test at once. When scoring a test, the response made by an examinee to each item is dichotomously scored. A correct response is given a score of 1, and an incorrect response a score of 0; the examinee's raw test score is obtained by adding up the item scores. This raw test score will always be an integer number and will range from 0 to N, the number of items in the test. If examinees were to take the test again, assuming they did not remember how they previously answered the items, a different raw test score would be obtained. Hypothetically, an examinee could take the test a great many times and obtain a variety of test scores. One would anticipate that these scores would cluster themselves around some average value. In measurement theory, this value is known as the true score and its definition depends upon the particular measurement theory. In item response theory, the definition of a true score according to D.N. Lawley is used. The formula for a true score is given in equation 4-1 below:

$$TS_j = \sum_{i=1}^{N} P_i(\theta_j) \qquad [4\text{-}1]$$

where: TS_j is the true score for examinees with ability level θ_j.
i denotes an item and $P_i(\theta_j)$ depends upon the particular item characteristic curve model employed.

The task at hand is to calculate the true score for those examinees having a given ability level. To illustrate this, the probability of correct response for each item in a four-item test will be calculated at an ability level of 1.0. This can be done using the formula for a two-parameter model and the procedures given in Chapter 2.

Item 1:

$P_1(1.0) = 1/(1 + EXP(-.5(1.0 - (-1.0)))) = .73$

To make the process a bit clearer, the dashed line on the figure below shows the relation between the value of θ and $P_1(\theta)$ on the item characteristic curve.

ITEM CHARACTERISTIC CURVE
TWO-PARAMETER   MODEL
B= -1   A= .5
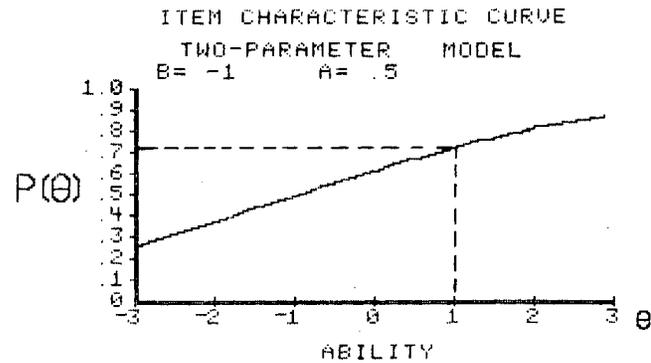
FIGURE 4-1. Item characteristic curve for item 1

Item 2:

$$P_2(1.0) = 1/(1 + \exp(-1.2\,(1.0 - (.75)))) = .57$$

ITEM CHARACTERISTIC CURVE
TWO-PARAMETER   MODEL
B= .75   A= 1.2
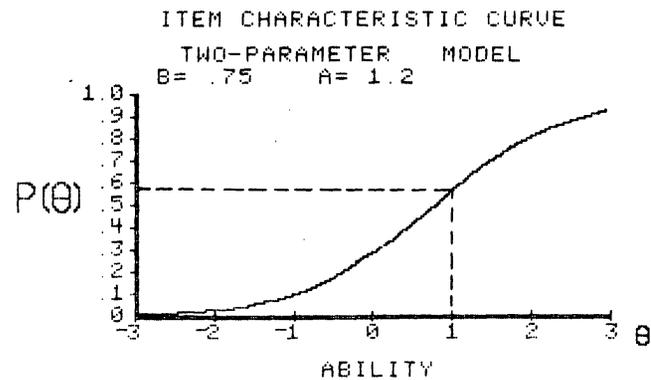
FIGURE 4-2. Item characteristic curve for item 2

Item 3:

$$P_3 (1.0) = 1/(1 + \text{EXP} (-.8 (1.0 - (0)))) = .69$$

ITEM CHARACTERISTIC CURVE
TWO-PARAMETER MODEL
B= 0   A= .8
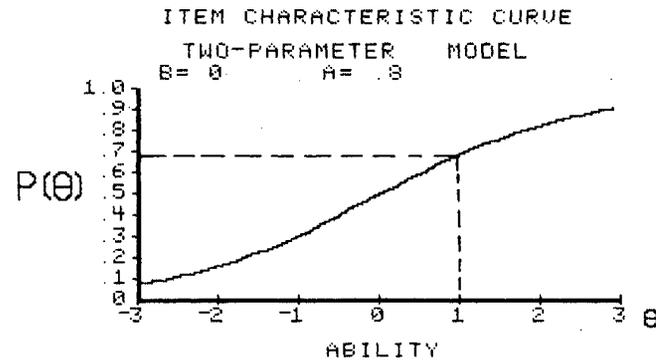
FIGURE 4-3. Item characteristic curve for item 3

Item 4:

$$P_4 (1.0) = 1/(1 + \text{EXP}(-1.0 (1.0 - (.5)))) = .62$$

ITEM CHARACTERISTIC CURVE
TWO-PARAMETER MODEL
B= .5   A= 1
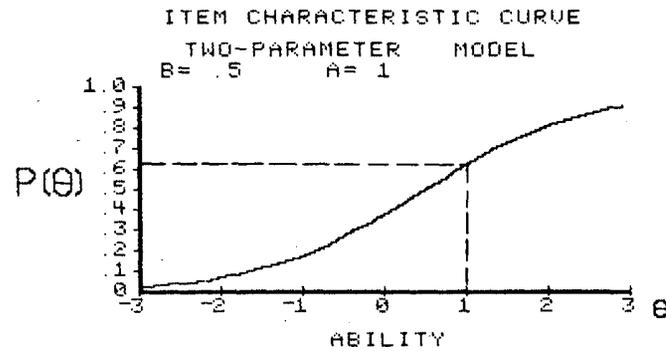
FIGURE 4-4. Item characteristic curve for item 4

Now, to get the true score at $\theta = 1.0$, the probabilities of correct response are summed over the four items:

TS = .73 + .57 + .69 + .62 = 2.61

Thus, examinees having an underlying ability of 1.0 would have a true score of 2.61 on this test. This score is intuitively reasonable because at an ability score of 1.0, each of the item characteristic curves is above .5 and the sum of the probabilities would be large. While no individual examinee would actually get this score, it is the theoretical average of all the raw test scores that examinees of ability 1.0 would get on this test of four items had they taken the test a large number of times. Actual tests would contain many more items than four, but the true score would be obtained in the same manner.

The calculations performed above were for a single point on the ability scale. This true score computation could be performed for any point along the ability scale from negative infinity to positive infinity. The corresponding true scores then could be plotted as a function of ability. The vertical axis would be the true scores and would range from zero to the number of items in the test. The horizontal axis would be the ability scale. These plotted scores would form a smooth curve, the test characteristic curve. The figure below depicts a typical test characteristic curve for a test containing 10 items.
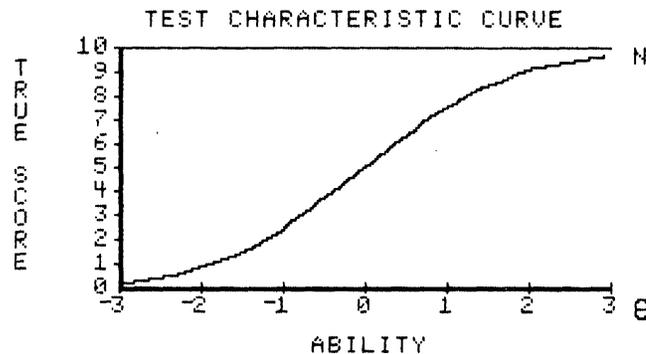
FIGURE 4-5. Test characteristic curve

The test characteristic curve is the functional relation between the true score and the ability scale. Given any ability level, the corresponding true score can be found via the test characteristic curve. For example, in Figure 4-5, draw a vertical line at an ability score of 1.0 upward until the test characteristic curve is intersected. Then, draw a horizontal line to the left until it intersects the true score scale. This line yields a true score of 7.8 for an ability score of 1.0.

When a one- or two-parameter model is used for the N items in a test, the left tail of the test characteristic curve approaches zero as the ability score approaches negative infinity; its upper tail approaches the number of items in the test as the ability score approaches positive infinity. The implication of this is that under these two models, a true score of zero corresponds to an ability score of negative infinity, and a true score of N corresponds to an ability level of positive infinity. When a three-parameter model is used for the N items in a test, the lower tail of the test characteristic curve approaches the sum of the guessing parameters for the test items rather than zero. This reflects the fact that under this model, very low-ability examinees can get a test score simply by guessing. The upper tail of the test characteristic curve still approaches the number of items in the test. Hence, a true score of N corresponds to an ability of positive infinity under all three-item characteristic curve models.

The primary role of the test characteristic curve in item response theory is to provide a means of transforming ability scores to true scores. This becomes of interest in practical situations where the user of the test may not be able to interpret an ability score. By transforming the ability score into a true score, the user is given a number that relates to the number of items in the test. This number is in a more familiar frame of reference and the user is able to interpret it. However, those familiar with item response theory, such as you, can interpret the ability score directly. The test characteristic curve also plays an important role in the procedures for equating tests.

The general form of the test characteristic curve is that of a monotonically increasing function. In some cases, it has a rather smooth S-shape similar to an item characteristic curve. In other cases, it will increase smoothly, then have a small plateau before increasing again. However, in all cases, it will be asymptotic to a value of N in the upper tail. The shape of the test characteristic curve depends upon a number of factors, including the number of items, the item characteristic curve model employed, and the values of the item parameters. Because of this, there is no explicit formula, other than

equation 4-1, for the test characteristic curve as there was for the item characteristic curve. The only way one can obtain the test characteristic curve is to evaluate the probability of correct response at each ability level for all the items in the test using a given item characteristic curve model. Once these probabilities are obtained, they are summed at each ability level. Then the sums are plotted to get the test characteristic curve. It is very important to understand that the form of the test characteristic curve does not depend upon the frequency distribution of the examinees' ability scores over the ability scale. In this respect, the test characteristic curve is similar to the item characteristic curve. Both are functional relations between two scales and do not depend upon the distribution of scores over the scales.

The test characteristic curve can be interpreted in roughly the same terms as was the item characteristic curve. The ability level corresponding to the mid-true score (the number of items in the test divided by 2, i.e., $N/2$) locates the test along the ability scale. The general slope of the test characteristic curve is related to how the value of the true score depends upon the ability level. In some situations, the test characteristic curve is nearly a straight line over much of the ability scale. In most tests, however, the test characteristic curve is nonlinear and the slope is only descriptive for a reduced range of ability levels. Since there is no explicit formula for the test characteristic curve, there are no parameters for the curve. The mid-true score defines the test difficulty in numerical terms, but the slope of the test characteristic curve is best defined in verbal terms. For most interpretive uses, these two descriptors are sufficient for discussing a test characteristic curve that has been plotted and can be visually inspected.

Computer Session for Chapter 4

This session has several purposes. The first is to show the form of the test characteristic curve and have you develop a feel for how true scores and ability are related in various tests. The second is to show the dependence of the form of the test characteristic curve upon the mix of item parameters occurring in the N items of the test. The computer session allows you to set the values of the item parameters for the N items of the test. The computer will plot the resultant test characteristic curve.

Procedures for an Example Case

This example will illustrate how to obtain a test characteristic curve for a small test.

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

b. Use the mouse to highlight the TEST CHARACTERISTIC CURVE session and click on [CONTINUE]. The TEST SPECIFICATION screen will appear.

c. Click on [NUMBER OF ITEMS] and set the number of items in the test to $N = 4$.

d. In the SELECT ITEM CHARACTERISTIC CURVE MODEL list, click on TWO PARAMETER.

e. In the SELECT ITEM PARAMETER CREATION METHOD list, click on USER INPUT OF ITEM PARAMETER VALUES.

f. Respond to the question SETTINGS OK? by clicking on the YES button. The ITEM PARAMETERS screen will appear.

g. Click on [ENTER PARAMETERS] and set the following item parameter values:

$b = -1.0,$ $a = .50$

$b =$ .75, $a = 1.2$

$b =$ 0.0, $a =$ .8

$b =$ .5, $a =$ .75

h.   Study the table of item parameters for a moment. If you need to change a value, click on the value and the data input box will appear, allowing you to enter a new value.

i.   When you are satisfied with the parameter values, respond to the message PARAMETER VALUES OK? by clicking on the YES button.

j.   Click on [CONTINUE]. The test characteristic curve shown below will appear on the screen.
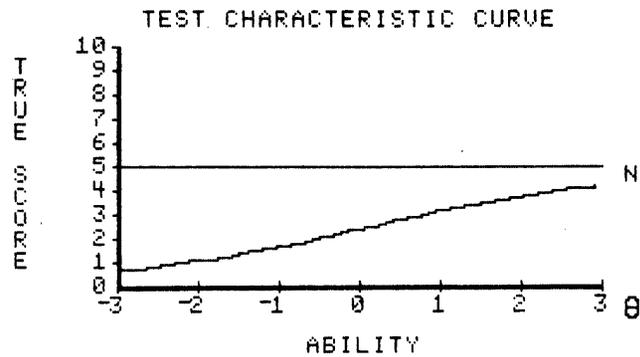
FIGURE 4-6. Test characteristic curve for the
example problem

In the display, a horizontal line has been drawn at the number of items in the test. This was done to remind you that the maximum value of a true score is the number of items in the test. This technique also allows the

computer program to use the same graph to show test characteristic curves based upon a different number of items.

It should be noted that because the ability range has been restricted arbitrarily to -3 to +3, the test characteristic curve may not get very close to either its theoretical upper or lower limits in the plotted curves. You should keep in mind that had the ability scale gone from negative infinity to positive infinity, the theoretical limits of the true scores would have been seen.

k. The test characteristic curve for this five-item test is very close to a straight line over the ability range from -2 to +2. Outside these values, it curves slightly. Thus, there is almost a linear relationship here between ability and true scores having a slope of about .5. The small slope reflects the low to moderate values of the item discrimination parameters. The mid-true score of 2.5 occurs at an ability level of zero, which reflects the average value of the b's.

l. Click on [CONTINUE]. The SELECT OPTION FROM LIST screen will appear.

m. A list will appear. If you click on MODIFY EXISTING TEST, the ITEM PARAMETERS screen will appear, allowing you to edit any of the values. If you click on CREATE A NEW TEST, the TEST SPECIFICATION screen will appear.

Exercises

a. Using a Rasch model

Exercise 1

(1) The TEST SPECIFICATION screen will appear.

(2) Click on [NUMBER OF ITEMS] and set the number of items in the test to $N = 5$.

(3) In the SELECT ITEM CHARACTERISTIC CURVE MODEL list, click on RASCH.

(4)   In the SELECT ITEM PARAMETER CREATION METHOD list, click on USER INPUT OF ITEM PARAMETER VALUES.

(5)   Respond to the question SETTINGS OK? by clicking on the YES button. The ITEM PARAMETERS screen will appear.

(6)   Click on [ENTER PARAMETERS] and set all the item difficulty parameters to a value of $b = 0.0$.

(7)   When you are satisfied with the parameter values, respond to the message PARAMETER VALUES OK? by clicking on the YES button.

(8)   Click on [CONTINUE]. The TEST CHARACTERISTIC CURVE screen will appear and the test characteristic curve will appear on the screen.

(9)   The test characteristic curve in this case looks just like an item characteristic curve for an item with $b = 0$. The mid-true score occurs at an ability level of zero.

(10)   Click on [CONTINUE]. The SELECT OPTION FROM LIST screen will appear. After you click on CREATE NEW TEST, the TEST SPECIFICATION screen will appear and you can do another exercise.

Exercise 2

(1)   Set $N = 10$

(2)   In the SELECT ITEM CHARACTERISTIC MODEL list, click on
       RASCH.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
       list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the difficulty parameters to:

       $b = -3$, $b = -2.5$, $b = -2.0$ $b = -1.5$, $b = -1.0$, $b = -.5$,
       $b = 0.0$, $b = .5$, $b = 1.0$, $b = 1.5$

(5)   The resulting test characteristic curve has a nearly linear section
       from an ability level of -3 to +1. After this point, it bends over
       slightly as it approaches a score of N. The mid-true score of 5
       corresponds to an ability level of -.6.

(6)   On the SELECT OPTION FROM LIST screen, click on CREATE
       NEW TEST. The TEST SPECIFICATION screen will appear and
       you can do the next exercise.

Exercise 3

(1)   Set $N = 10$.

(2)   Select a Rasch item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
       list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item difficulty parameters to:

$b = -.8$, $b = -.5$, $b = -.5$, $b = 0$, $b = 0$, $b = 0$, $b = .5$, $b = .5$,
$b = .5$, $b = .8$

(5)   The test characteristic curve has a well-defined S-shape much like
        an item characteristic curve. Only the section near an ability level of
        zero is linear. The mid-true score of 2.5 corresponds to an ability
        score of .05.

(6)   Click on [CONTINUE]. The SELECT OPTION FROM LIST
        screen will appear. Click on CREATE NEW TEST. The TEST
        SPECIFICATION screen will appear and you can do another
        exercise.

b. Using a two-parameter model

Exercise 1

(1)   Set $N = 5$.

(2)   Select a two-parameter item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
        list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameters to:

$b = 0$, $a = .4$
$b = 0$, $a = .8$
$b = 0$, $a = .4$
$b = 0$, $a = .8$
$b = 0$, $a = .4$

(5)   The test characteristic curve is nearly a straight line with a rather
        shallow slope reflecting the low to moderate values of *a*. The mid-
        true score of 2.5 occurs, as expected, at an ability level of 0.

Exercise 2

(1)   Set $N = 5$.

(2)   Select a two-parameter item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
       list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameters to:

   $b = 1, a = 1.6$
   $b = 1, a = 1.9$
   $b = 1, a = 1.6$
   $b = 1, a = 1.9$
   $b = 1, a = 1.6$

(5)   The majority of the test characteristic curve is compressed into a
       rather small section of the ability scale. Up to an ability level of -1,
       the true score is nearly zero. Beyond an ability level of 2.5, the true
       score is approaching a value of N. Between these two limits, the
       curve has a definite S-shape, and the steep slope reflects the high
       values of the discrimination parameters. The mid-true score of 2.5
       occurs at an ability level of 1.0. Notice how the difference in
       average level of discrimination in these last two problems shows in
       the difference of the steepness of the two curves.

Exercise 3

(1)   Set $N = 5$.

(2)   Select a two-parameter item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
       list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameter values to:

   $b = -2.0,$ $a = .4$
   $b = -1.5,$ $a = 1.7$

$$b = -1.0, \qquad a = .9$$
$$b = -.5, \qquad a = 1.6$$
$$b = 0.0, \qquad a = .8$$

(5)   The test characteristic curve has a moderate S-shape and has a mid-true score of 2.5 at an ability level of -.8, which is not the average value of b but is close to it.

c. Using a three-parameter model

Exercise 1

(1)   Set $N = 5$.

(2)   Select a three-parameter item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameters to:

$$b = 1.0, \qquad a = 1.2, \qquad c = .25$$
$$b = 1.2, \qquad a = .9, \qquad c = .20$$
$$b = 1.5, \qquad a = 1.0, \qquad c = .25$$
$$b = 1.8, \qquad a = 1.5, \qquad c = .20$$
$$b = 2.0, \qquad a = .6, \qquad c = .30$$

(5)   The test characteristic curve has a very long lower tail that levels out just above a true score of 1.2, which is the sum of the values of the parameter c for the five items. Because of the long lower tail, there is very little change in true scores from an ability level of -3 to 0. Above an ability level of zero, the curve slopes up and begins to approach a true score of N. The mid-true score of 2.5 corresponds to an ability level of about .5. Thus, the test functions among the high-ability examinees even though, due to guessing, low-ability examinees have true scores just above 1.2.

Exercise 2

(1)   Set $N = 10$.

(2)   Select a three-parameter item characteristic curve model.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
        list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameters to:

$b =$ 2.34, $a = 1.9$, $c = .3$
$b =$ -1.09, $a = 1.64$, $c = .3$
$b =$ -1.65, $a = 2.27$, $c = .07$
$b =$ - .40, $a =$ .94, $c = .12$
$b =$ 2.90, $a = 1.83$, $c = .16$
$b = $-1.54, $a = 2.67$, $c = .27$
$b = $-1.52, $a = 2.01$, $c = .17$
$b = $-1.81, $a = 1.98$, $c = .27$
$b =$ - .63, $a =$ .92, $c = .28$
$b = $-2.45, $a = 2.54$, $c = .07$

(5)   Compared to the previous test characteristic curves, this one is quite
        different. The curve goes up sharply from an ability level of -3 to -1.
        Then it changes quite rapidly into a rather flat line that slowly
        approaches a value of N. The mid-true score of 5.0 is at an ability
        level of -1.5, indicating that the test functions among the low-ability
        examinees. This reflects the fact that all but two of the items had a
        negative value of the parameter b.

d. Exploratory exercises

(1)   Set $N = 10$.

(2)   Select an item characteristic curve model of your choice.

(3)   In the SELECT ITEM PARAMETER CREATION METHOD
        list, click on USER INPUT OF ITEM PARAMETER VALUES.

(4)   Set the item parameters to values of your choice.

(5)   When the test characteristic curve appears on the screen, make a
        sketch of it on a piece of paper.

(6)   When the SELECT OPTION FROM LIST screen appears,
        respond to the message MODIFY EXISTING TEST? by typing
        YES.

(7)   Change some of the values of the item parameters and then display
        the new test characteristic curve. Place this curve on the same graph
        as the previous curve.

(8)   Repeat the editing process until you can predict what effect the
        changed item parameters will have on the form of the test
        characteristic curve.

(9)   To explore the effect of modifications to sets of item parameters,
        return to Step 3 and click on COMPUTER GENERATED ITEM
        PARAMETER VALUES.

(10)   Respond to SETTINGS OK? by selecting YES.

(11)   Click on [GENERATE ITEM PARAMETER VALUES].

(12)   Respond to PARAMETER VALUES OK? by selecting YES. Try
        to estimate the form the test characteristic curve will take, then click
        on [CONTINUE].

(13)   After viewing the curve, press [CONTINUE] and then experiment
        with further modifications to the existing test.

Things To Notice

1. Relation of the true score and the ability level.

    a. Given an ability level, the corresponding true score can be found via the test characteristic curve.

    b. Given a true score, the corresponding ability level can be found via the test characteristic curve.

    c. Both the true scores and ability are continuous variables.

2. Shape of the test characteristic curve.

    a. When $N = 1$, the true score ranges from 0 to 1 and the shape of the test characteristic curve is identical to that of the item characteristic curve for the single item.

    b. The test characteristic curve does not always look like an item characteristic curve. It can have regions of varying steepness and plateaus. Such curves reflect a mixture of item parameter values having a large range of values.

    c. The ability level at which the mid-true score ($N/2$) occurs depends primarily upon the average value of the item difficulty parameters and is an indicator of where the test functions on the ability scale.

    d. When the values of the item difficulties have a limited range, the steepness of the test characteristic curve depends primarily upon the average value of the item discrimination parameters. When the values of the item difficulties are spread widely over the ability scale, the steepness of the test characteristic curve will be reduced, even though the values of the item discriminations remain the same.

    e. Under a three-parameter model, the lower limit of the true scores is the sum of the values of the parameter c for the N items of the test.

    f. The shape of the test characteristic curve depends upon the number of items in the test, the item characteristic curve model, and the mix

of values of the item parameters possessed by the N items in the test.

3. It would be possible to construct a test characteristic curve that decreases as ability increases. To do so would require items with negative discrimination for the correct response to the items. Such a test would not be considered a good test because the higher an examinee's ability level, the lower the score expected for the examinee.