

The One-Parameter Logistic Model (1PLM) And Its Application In Test Development

Mfonobong E. Umobong

Professor, Dept. of Educational Foundations
University of Uyo, Uyo

ABSTRACT

Recent development in the measurement field has led to the development of various measures for the purpose of constructing and validating assessment instruments. One of such is the most popular 'item response theory (IRT)'. Various models have been developed for the IRT of which the one parameter logistic model (1PLM) is part. This paper gives a brief overview of the 1PLM focusing on explaining the theoretical perspectives, the assumptions underlying it, the item characteristic curve as well as an illustrative example in the calculation of the 1PLM. The model promises to be one effective way of examining the relationship between examinees ability and his/her performance on a given task and a potential guide for test development.

INTRODUCTION AND BACKGROUND

Item response theory (IRT) has become a very popular topic in the measurement field. It is being used by many test publishers, departments of education, in industrial and professional organizations for the purpose of testing.

Item Response Theory (IRT) is a modeling technique that tries to describe the relationship between an examinee's test performance and the latent trait underlying the performance (Hernard, 2000). This theory postulates that (a) an examinee test performance can be explained by a set of factors called traits, latent traits, or abilities and (b) the relationship between examinee test performance and the set of traits assumed to be influencing test performance can be described by a monotonically increasing function called an item characteristics function (Hambleton, 1982).

IRT assumes that there is a single ability or dominant factor that explains performance, this ability parameter is often denoted as θ . In IRT, item parameter estimates are generated and used to describe the test items, while ability estimates are obtained to describe the performance of the examinees. Item response models are known to generate invariant item statistics and ability estimates such that the item parameter estimates are not dependent on the characteristics of the examinees and the ability estimates are not dependent on the items. The two desirable features of IRT are obtainable when an item response model fits a test data. Item response theory employs three different models viz. the one -, two -, and three parameter model. The purpose of this paper is to give a brief discussion of the one parameter model. The focus will be on explaining the theoretical perspective of one parameter logistic model (IPLM), its assumptions, item characteristic curve and an illustrative example of the calculation of IPLM will also be shown.

THEORETICAL MODEL DEVELOPMENT

The item response model was built based on the normal ogive, which is a standardized form of an ogive. An ogive or a cumulative frequency polygon, or cumulative percentage is the graph of

cumulative frequency distribution which is particularly useful in determining the various percentile point in a distribution of scores (Hinkle, Wiersma and Jurs, 1998). The normal ogive also called the normal frequency function (f.f) even when the ordinate (vertical axis) is defined as frequency, proportion, percent or density has the formula:

$$\int N(0,1) \text{ or } \int_{-\infty}^Z N(0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-\frac{z^2}{2}} dz \quad 1.1$$

The definition has an integral sign \int on the right side which means that no algebraic function can be found to describe the normal ogive. This makes working with the normal curve very cumbersome mathematically and thus require numerical method to solve, or a table of values. The tediousness usually experienced in the calculation of the normal ogive for item response theory (IRT) modeling has led to the use of logistic frequency function as a complimentary procedure to the normal ogive (Crocker and Algina, 1986). The logistic frequency function used also has a mean of zero and a standard deviation of 1.0. The formula for this logistic frequency function (l.f.f.) is given as:

$$\int_{-\infty}^Z L(0,1.7) = \frac{1}{1+e^{-1.7z}} \quad 1.2$$

The 1.7 in the exponent is chosen to allow the logistic frequency function to approximate the normal frequency function as closely as possible, and in some cases represented by the upper case letter D. The letter e is the base of natural logarithms: $e \approx 2.118281828$. Since the logistic ogive has no integral sign in its definition, it is very easy to work with and usually substituted as a convenient and very close approximation of the normal ogive.

The logistic item response model (LIRM) is a monotonic S curve, increasing from left to right, always gets higher and higher, never completely horizontal and never gets down. The LIRM has a lower and upper asymptote indicating that it will never be equals to 1 or 0 at any point in time. An example of a normal ogive is illustrated in figure 1.

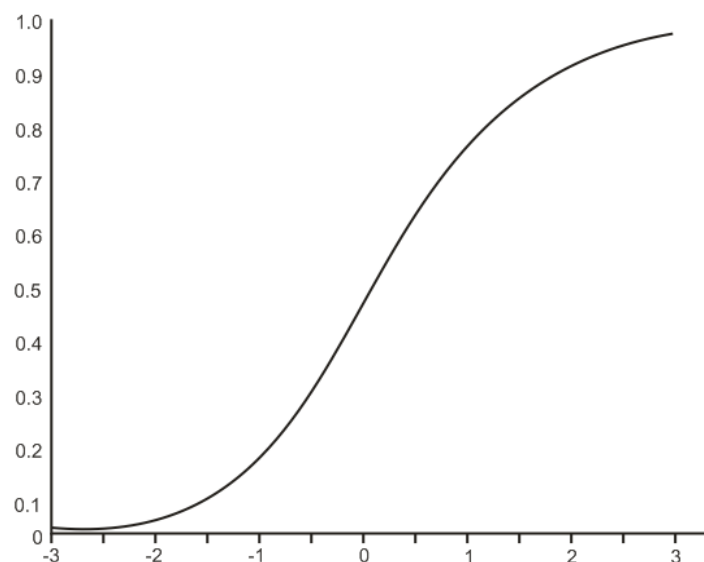


Figure 1: Normal ogive

As seen in this figure, between -2.0 and -0.5 on the horizontal axis, the ogive is concave upward and between 0.5 and 2.0 it is concave downward. At a point between -0.5 and 0.5 the ogive changes from being concave upward to being concave downward and that point is called

“inflection point” (Baker, 2001). The inflection point is the point where the slope of the ogive is at its maximum and for this ogive it is located at 0.50 on the vertical axis and 0.0 on the horizontal axis. In IRT, the horizontal position of the inflection point is called the “b-parameter” also called the item difficulty parameter or threshold and represents the point on the ability scale \emptyset (the horizontal axis), where an individual has 50 percent probability of answering correctly the item. The inflection point is always halfway between the lower and upper asymptotes (Baker, 2001; Courville, 2004).

Logistic item response models are simply a form of logistic regression and Courville (2004) explained that the theory behind logistics regression is that if the dependent variable is a set of dichotomized scores, then it is possible to set the probability of a particular score. Thus with a single independent variable, the probability equation is given as:

$$P(\emptyset) = \frac{e^{B_0+B_i\emptyset}}{1+ e^{B_0+B_i\emptyset}} \quad 1.3$$

This can be represented as:

$$P(\emptyset) = \frac{1}{1+ e^{-(B_0+B_i\emptyset)}} \quad 1.4$$

Where B_0 is the y intercept and B_i is the slope of the function derived by a mathematical relationship between the independent and the dependent variable, e represents the base of the natural logarithm approximated as 2.178 (NORUSIS, 1990). The above equation can be extended to include multiple independent variables depending on the model that is used. The focus of IRT on item level information can be derived from the logistic equation. Since the concern is with the ability scale called \emptyset , then the horizontal axis can be represented by \emptyset and substituted for z in the normal ogive or l.f.f. The b-parameter can also be included by estimating it from the horizontal axis variable. The \emptyset parameter ranges from -3 to +3. The relationship between the latent ability parameter (\emptyset) and the probability of a particular score is a nonlinear function.

For the one-parameter model, the function is often given as:

$$P(\emptyset_i) = \frac{e^{Da(\emptyset-b_i)}}{1+ e^{Da(\emptyset-b_i)}} \quad 1.5$$

A look at this function will show its similarity with the single parameter logistic regression, with the D representing a constant usually set to 1.7. The b_i represent the item difficulty, the \emptyset parameter represents the level of the ability or latent trait measured for examinee. The higher the \emptyset values, the higher the examinee ability level measured by the item (Crocker and Algina, 1986).

For the one-parameter model the population distribution of the underlying ability or latent trait are conceived as being normally distributed in the population with a mean of zero and variance of one. The b parameter or item difficulty also called threshold helps in determining the position of the logistic curve along the ability scale. The further the curve is to the right, the more difficult the item will be. The ‘a’ parameter is the slope or item discrimination parameter representing the degree to which item response varies with ability \emptyset . For the one-parameter model, the item discrimination parameter is assumed to be constant. The one parameter model is often referred to as the Rasch model and is sometimes represented as:

$$P(Y_i = 1/\emptyset_i) = \frac{1}{1+ e^{[-a(\emptyset-b_i)]}} \quad 1.6$$

Which expresses that the probability of a correct ($Y_i = 1$) response to an item i is a function of ability of the examinee (θ), the threshold or difficulty parameter b_i and the discrimination parameter a . the slope as mentioned earlier is fixed at 1 for all items.

CHARACTERIZATION OF ONE PARAMETER LOGISTIC (IPL) MODEL

The mathematical models employed in IRT assume that an examinee's probability of answering correctly a given item depends on the examinees ability and the characteristics of the item. Examinees ability is considered the major characteristics of the person and is denoted as θ , it is also called the ability parameter. The ability parameter is conceived of as an underlying, unobservable latent construct or trait that helps an individual to perform or answer correctly an item. Also embedded in this model is the characteristic of the item also known as the item parameters. IRT is regarded as an improvement in measurement due to its ability to generate item parameters that are invariant.

Basically, there are three characteristics of an item response model but the most commonly used IRT models which is the 1 parameter is built off a single item parameter giving it the name the 1 - parameter model. This single parameter is the item difficulty, also referred to as the threshold parameter. Item difficulty measures the location of an item on the continuum. The item parameter is believed to be a continuum with the upper end indicating greater proficiency in whatever is measured than the lower end. This means that items located towards the right side of the continuum demands an individual to possess greater proficiency (ability) in order to answer correctly, than items located towards the left side of the continuum. It is also important that for any instrument to be regarded as good, the item difficulty parameter usually denoted as b should be located throughout the continuum with some above and others below 0.

Hambleton and Swaminathan (1985) explained that the characteristics of an item response model (IRM) involve four ideas:

- An IRT model must specify the relationship between the observed response and the underlying unobservable construct;
- The model must specify ways of estimating scores on the ability;
- The underlying unobservable construct can be estimated based on the examinee's scores
- An IRT model assumes that the performance of an examinee can be predicted or explained completely from one or more abilities.

These four characteristics of an IRT model may at first glance be similar to the classical test (CT) model but there are major differences that seem to set IRT apart. For instance, IRT models are noted for their ability to generate estimates that are invariant. Such that the item parameter estimates (item difficulty) are said to be "person-free" and not dependent on the characteristics of the examinee. This means that the item difficulty statistics for instance would not change if different persons were used. Also, the ability estimates are not dependent on the items but are said to be "item free" meaning it would not change if different items were used. This is the underlying basis of the 1-parameter model that allows for objectivity in measurement.

Assumptions Underlying the One-Parameter Model

The one-parameter IRT model is based on two basic assumptions. These are unidimensionality and local independence.

Unidimensionality assumption which is the most common assumption of all IRT model states that one and only one ability is measured by a set of items in a test. Unidimensionality does not mean that the items correlate positively with each other rather it is conceived that all items should be negatively correlated with each other and still be unidimensional (de Ayala, 2001). This is the most restrictive and complex assumption of IRT which sometimes can never be met. However, what is required is that a single dominant factor or component should underlie item responses. This single dominant factor is usually the ability measured by the test. In a test measuring mathematics proficiency, for an examinee to answer item correctly, he must possess the mathematical skills (ability) required by the test. however if the items in the test measure some verbal ability then a new ability (verbal) is introduced and such a test is no longer unidimensional rather, it is referred to as a multidimensional test because more than one ability is necessary to account for examinee test performance. The assumptions of unidimensionality of a test can be assessed by different methods, the most popular is the factor analytic approach.

The second and related assumption is **local independence**. Local independence assumes that if the ability is held constant, examinee's responses to any pair of items are statistically independent. This means that the probability of an examinee responding to a test item is not affected by the answers given to other items in the test. The item responses are assumed to be independent of one another with the only relationship among the items being explained by the conditional relationship with ability θ . In other words, it is the abilities specified by the model that is the only factor influencing examinees responses to items in a test. A violation of this assumption may result in parameter estimates that are different from what would have been expected if the data were locally independent. The assumptions of unidimensionality and local independence are related in that items found to be locally dependent will appear as a separate dimension in the factor analysis.

Item Characteristic Curve or Item Response Function

All IRT models are derived to generate item characteristic curves, because the basic concepts of IRT rest on the individual items that make up a test and not on the aggregate of item responses which is usually the test score. The interest in IRT is on whether an examinee got each item correct or not. One basic principle in IRT is that each examinee responding to a test item possesses some amount of an underlying ability. Therefore, the examinee is expected to have a numerical value that will place him/her on the ability scale. This underlying ability or latent trait is usually denoted as θ . It is believed that at each ability level, an examinee with that ability will have a certain probability of a given correct answer to an item. This probability can be denoted as $P(\theta)$, and will be high for examinee that have a high ability and low for examinee with low ability. If one is to plot the probabilities ($P(\theta)$) for many examinees that responded to the item, as a function of the amount of ability and the points is connected, the result will be an S-shaped curve as shown in figure 2. The S-shaped curve shows that the probability of a correct response is near zero at the lowest levels of ability and increases until at the highest level of ability as the probability of correct response approaches 1. The S-shaped curve describes the relationship between the probability of a correct response to an item and the ability or latent trait scale. This curve is called the item response function (IRF) and until recently was called item characteristic curve (ICC). These terms IRF and ICC are used interchanging in the literature and for this paper, the ICC is used to denote the functional relationship. In any test, each item will have its own ICC.

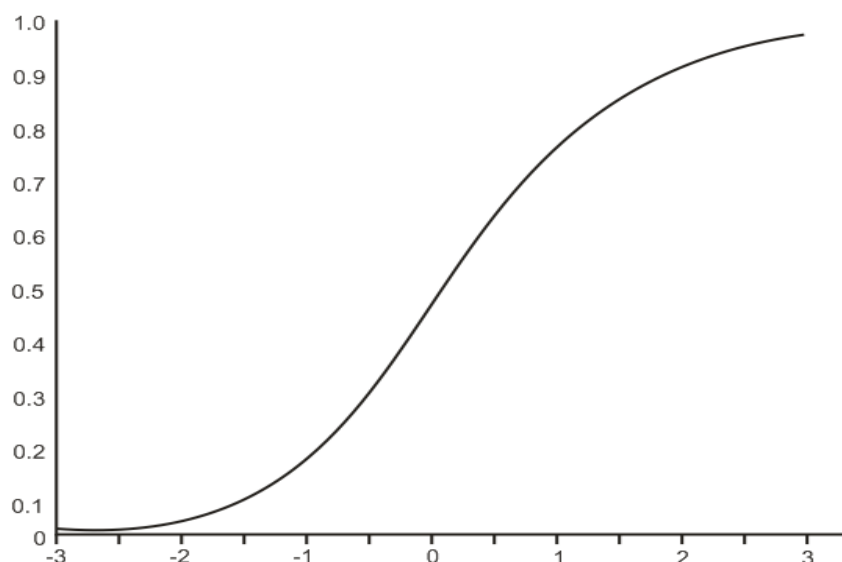


Figure 2: A typical item Characteristic Curve

The ICC shows a near ideal case and indicates that when θ is zero that is average, then the probability of answering an item correct is almost 0.5. This shows that at this point, an individual has a 50:50 chance of correctly responding to the item. As the θ increases, there is an increase probability that the examinee will correctly respond to an item. An item characteristic curve plots the probability that an examinee will respond correctly to an item as a function of a tests' latent trait. Crocker & Algina (1986) explained that two interpretations can be discerned from IRT, (i) for a correct response, there is the probability that a randomly chosen member of homogeneous subpopulation will response correctly to an item, and (ii) that the probability represents the probability of a specific examinee responding correctly for a subpopulation of items.

The item response function or ICC is the building block of IRT as all the other constructs depends on this curve (Baker, 2001). There are some features of ICC that makes it distinct such as; that ICC cannot be higher or never reaches 1.0; the curve rises rapidly as we move from left to right and is said to be strictly monotonic; the curve has a lower asymptote of 0.0. (See Figure 2). It can be noticed that these features are exactly what is obtained with the normal or logistic ogive, meaning that the normal or logistic ogive or function can be used to describe ICC very well. The logistic function provides a standard mathematical model for the ICC which defines a family of curves as the general shape of the ICC.

As mentioned earlier, ICC shows the probability ($P(\theta)$) of a correct response as a function of the ability θ . It should be noted that the probability of a correct response depends not only on the ability but also on the properties of the item, also known as the item parameters. In IRT, models have been developed as simple as using one parameter or as complex as two or three parameters namely a, b or c parameters. The simplest IRT model for a dichotomously scored item is that having one item parameter. The ICC which is the probability of a correct response given the single item parameter, b_i as a function of the individual latent trait (ability) level θ can be represented as shown in figure 2. The function shown in the figure is usually referred to as the one-parameter logistic function or in IRT as a one parameter model with its values ranging between 0 and 1. The theoretical range for the item parameter is from $-\infty$ to $+\infty$ but typically the item and person parameter ranges between -3 to +3.

From the figure, it can be seen that both items and persons are located at the same continuum in the horizontal axis, reason being that in IRT, the ability of a person is usually equated with the difficulty of a test problem. This is to allow for a comparative statement about how a typical examinee with a particular ability might response to an item. George Rasch, a Danish mathematician in 1960 developed the one-parameter logistic model. The ICC for the one - parameter logistic model is given by the equation 1.7 below:

$$P(\emptyset) = \frac{1}{1 + e^{-1(\emptyset - b_i)}} \quad 1.7$$

or

$$P_i(\emptyset) = \frac{e^{(\emptyset - b_i)}}{1 + e^{-(\emptyset - b_i)}} \quad 1.8$$

Where:

b_i is the difficulty parameter

\emptyset is the ability parameter

$P_i\emptyset$ is the probability that a randomly selected examinee with ability \emptyset will answer correctly an item i

e is an exponential number, whose value is 2.718.

In the one parameter logistic model, the probability of a correct response is as a result of the interaction between the examinees ability \emptyset and the item parameter b_i . The one-parameter model assumes that for all items in the test, the discrimination index will be constant and thus, the value may be set at 1. The b_i parameter which is the item difficulty is also conceived of as the location parameter. Item difficulty is a location index that indicates the position of the ICC or where the item functions along the ability scale (de Ayala, 2009). It is an indication of how easy or difficult an item is. The higher the b_i values (parameter) the higher the ability required for an examinee to have a 50% chance of getting an item correctly, hence the more difficult the items. Difficult items are located to the right or to the higher end of the ability scale while easier items are located to the lower end or left of the ability scale.

To explain further, we refer back to figure 2 since the lower end (to the left of the continuum) represents less ability for example in a test of mathematics than the upper end, then items located in the lower end require less ability to be correctly answered than those in the upper (right) end. Those items located at the upper end of the continuum require an examinee to have greater ability to get the items correct. Using the typical values of b_i from -3 to +3, then items whose b values are near -3 will correspond to items that are very easy while those with values near +3 will corresponds to items that are very difficult for the examinees.

In the I-parameter model, all items in the test have ICC that is the same shape with the only distinguishing characteristics of an item from another being the left-right location of the ICC on the horizontal axis of the ability scale \emptyset . Figure 3 presents three items characteristic curve or item response functions on the same graph.

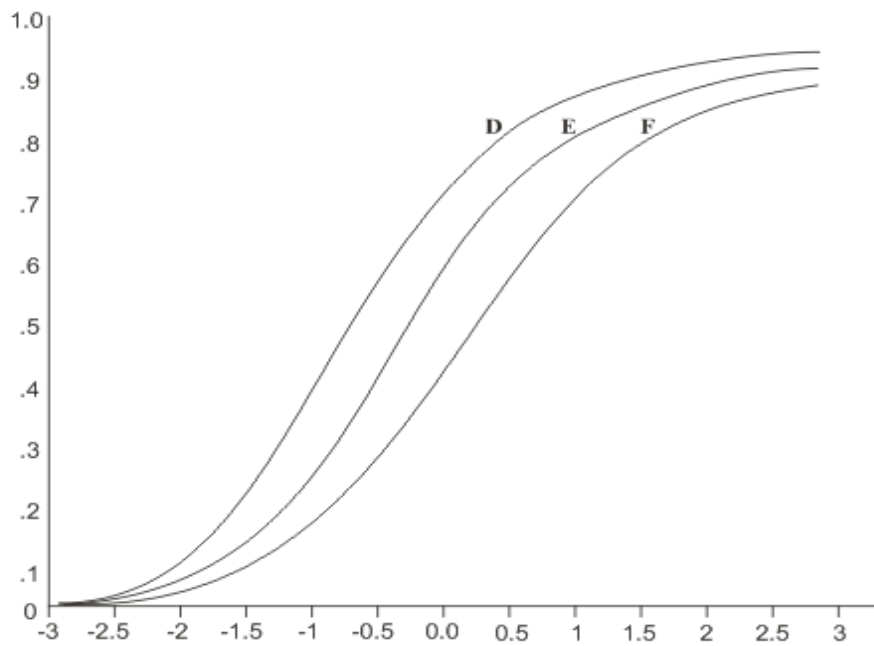


Figure 3: One parameter ICC for three items D, E and F

The three ICC representing items D, E & F all have the same level of discrimination but differs only on their location on the ability scale. In the IPLM, it is assumed that the only characteristic that influences examinee performance is the item difficulty (b_i) and all items are equally discriminating. The item parameter b -values for items D, E, F are -0.5, 0.0 and 1.0 respectively. Item D is an easy item with the low ability examinees having the probability of correctly responding to the item while the probability for the high ability examinee approaches 1. Item E represents an item of medium difficulty such that the probability of a correct response is low at the lowest ability levels and near 1 at the highest ability levels. Item F represents a hard item with the probability of correctly responding being low for most of the ability scale and only increasing at the higher ability levels.

It is also noticed from figure 3 that the lower asymptote for all the three ICC is zero, meaning that examinees of very low ability have zero probability of answering correctly the items. Here, no allowance is made for the fact that examinees of low ability have the likelihood of guessing correctly. The one-parameter model is therefore based on restrictive assumptions.

ILLUSTRATIVE EXAMPLE

To illustrate how the one-parameter logistic model is used to compute the point and plot an ICC. Consider an item with a b value of 1.5, and compute the probability at $\theta = -3$ to $+3$ at 0.5 intervals. The equation given in 1.7 is used and since for the one parameter model the discrimination index is constant, the equation is therefore represented as:

$$P(\theta_i) = \frac{1}{1 + e^{-(\theta - b_i)}} \tag{1.9}$$

This can also be substituted as $P(\theta_i) = \frac{1}{1 + e^{-L}}$ where the logit $L = \theta - b_i$

The first term to be computed is the logit:

$$L = (\theta - b_i)$$

Substituting the value for $\theta = -3$, we have

$$L = -3 - 1.5 = -4.5$$

Next compute the exponential i.e

$$\text{Exp} (-L) = \text{exp} (-4.5) = 90.017$$

The denominator of the equation is computed as:

$$1 + \text{Exp} (-L) = 1.0 + 90.017 = 91.017$$

Finally, the value of p \emptyset can be obtained as:

$$P(\emptyset_i) = \frac{1}{1 + e^{-L}} = \frac{1}{91.017} = 0.011 \approx 0.01$$

For a $\emptyset = 3$, this can be calculated as: $L = a (\emptyset - b)$

Substituting the value, we have:

$$L = 3 - 1.5 = 1.5.$$

Next the exponential is computed $\text{Exp} (-L) = 0.223$

The denominator of the equation 1.7 is computed as:

$$1 + \text{Exp}(-L) = 1.22$$

Then the value of $P\emptyset$ can be obtained as:

$$P(\emptyset_i) = \frac{1}{1 + e^{-L}} = \frac{1}{1.22} = 0.8197 \approx 0.82$$

In Table 1 the calculation for the different ability levels using $b = 1$ and -1.5 is shown.

Table 1: Calculations for the One-parameter Model with $b = 1.0$ and -1.5

For $b_i = 1.0$					For $b_i = -1.5$				
\emptyset	LOGIT	EXP (-L)	1 + EXP (-L)	P	\emptyset	LOGIT	EXP (-L)	1 + EXP (-L)	P
-3	-4	54.598	55.598	.02	-3	-1.5	4.482	5.482	.18
-2.5	-3.5	33.115	34.115	.03	-2.5	-1.0	2.718	3.718	.27
-2	-3	20.086	21.086	.05	-2	-0.5	1.649	2.649	.38
-1.5	-2.5	12.182	13.182	.08	-1.5	0	1	2	.5
-1	-2	7.389	8.389	.12	-1	0.5	.607	1.607	.62
-0.5	-1.5	4.482	5.482	.18	-0.5	1.0	.368	1.368	.73
0	-1	2.718	3.718	.27	0	1.5	.223	1.223	.82
0.5	-0.5	1.649	2.649	.38	0.5	2	.135	1.135	.88
1	0	1	2	.50	1	2.5	.082	1.082	.92
1.5	0.5	.607	1.607	.62	1.5	3	.05	1.05	.95
2	1	.368	1.368	.73	2	3.5	.03	1.03	.97
2.5	1.5	.223	1.223	.82	2.5	4	.018	1.018	.98
3	2	.135	1.135	.88	3	4.5	.011	1.011	.99

From the above, it can be seen that it is very easy to compute the probability of correctly responding to an item at a given ability level using the logistic model. The corresponding item characteristic curve for the item in table 1 with $b = 1.0$ is shown as fig. 4

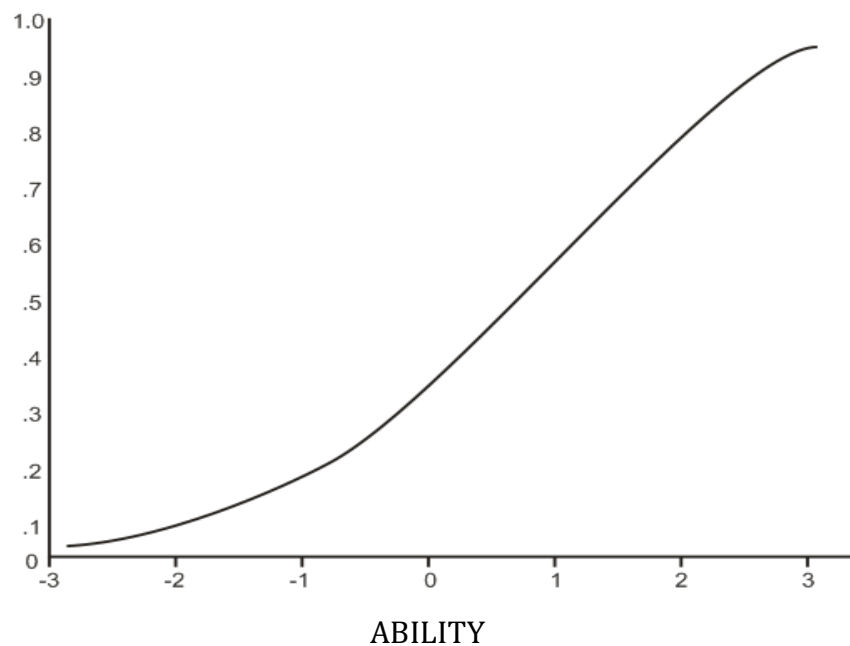


Figure 4: Item Characteristic Curve (ICC) for a one-parameter model

APPLICATION IN TEST DEVELOPMENT

As it has been shown, the one parameter logistic model is useful in test design most especially in item analysis, item selection, item banking, test equating and item bias or differential item functioning. According to Schumacker (2005) IRT models offer the following to test developers:

- ❖ Item statistics that are independent of the sample from which they are estimated.
- ❖ Examinee scores are independent of test difficulty.
- ❖ Item analysis accommodates matching test score to examinee knowledge level.
- ❖ Test analysis does not require strict parallel test for accessing reliability.
- ❖ Item statistics and examinee ability are both reported on the same scale.

In test development, the final sets of items are usually selected through the process of item analysis. In the 1PL model, item analysis usually involves the determination of the item difficulty. It is believed that item difficulty can be obtained that is sample dependent such that we can accumulate difficulty statistics for items over multiple samples of people.

With the one-parameter model, it is possible for test developers to have a better method of item selection. Usually, items are selected based on the amount of information they contribute to the overall amount of information in the test. Thus with the item information and item response function, and the test information function derived from the 1-parameter model, it is possible to select items that will provide the required amount of information needed. Using the 1-parameter model allows test developers to estimate the contribution of each item to the test information function independently of other items in the test (Hambleton & Jones, 1993).

In many testing programmes, two or more parallel forms of a test are administered to a group of examinee in the process of selection, certification or admission. To be sure that the scores from the different forms are comparable, test score equating are usually carried. The 1-parameter model believes that the ability (θ) of an examinee is invariant across different sets of items and therefore scores obtained from different tests can be scaled and compared appropriately. Thus the 1-parameter model provides a framework for solving many problems in measurement as a result, test developers and publishers can use the model in developing

tests, equating scores from different tests and reporting scores. This is especially so because the 1-parameter model allows the test developer to develop test that has the desired objectivity in measurement at nearly any defined ability level.

The 1-parameter model also has applicability in the development of item banks. Item banking has to do with the process of storing items for future use. Since the 1-parameter model allows for the estimation of item parameters that are sample independent, then it is possible to estimate parameters for certain items during field testing and use it for later testing. Since the 1-parameter model provides information needed to identify the strengths and weaknesses of the examinee, as the items are scaled to different ability level it is easier to build item banks. These banks will ultimately be used for computerized Adaptive Tests (CAT) and are very important especially for large scale testing programmes.

In test development, there is need to develop items that are fair to all the subgroups of examinee. Items that favour a group against the other are said to be biased and therefore does not provide the needed information, test developers therefore has to take into consideration the issue of test fairness. The one-parameter model provides a good framework for combating test bias through the process of determining Differential Item Functioning (DIF) in a test. The 1-parameter model or the Rasch Model uses 'Item Map' to model the item location with the estimate of ability.

CONCLUSION

The one-parameter logistic model (1PLM) is usually conceived of as a statistical approach of trying to model response data. The 1PL model assumes that all items in a test have a constant discriminating parameter (a) with the only distinguishing feature being the point of inflection or the difficulty index of the item. Thus the 1PLM derives its name from the fact that only one parameter of an item being the difficulty parameter is seen as accounting for the differences in individuals' ability.

For the IRT, both persons and items are located on the same continuum and for the 1PLM, the sum of the person's item responses (total test score) is a sufficient statistic for estimating his or her location (ability parameter) and the sum of the responses to an item is a sufficient statistic for estimating the item's location (item difficulty). In the 1PLM, the graphical representation of the relationship between an individual response on an item and the probabilities of a correct response to the item is called the item characteristic curve (ICC). The 1PLM can be used to compute the points and plot the ICC for each item in the test.

An illustrative example of this has been shown in this paper, however, recent computer programs have been developed such as the BILOG, WINSTEP, RASCAL and XCALIBRE that can plot the ICC for all items included in a test. It is therefore highly recommended that the 1PLM should be applicable in test development as it makes measurement more objective.

References

- Baker, F. B. (2001). *The basics of item response theory*. Washington DC: ERIC Clearing House.
- Courville, T. G. (2004). An empirical comparison of item response theory and classical test theory item/test statistics. *A Dissertation Submitted to Graduate School, Texas A & M University*.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- de Ayala, R. J. (2009). *The theory and Practice of item response theory*. New York: The Guildford Press.

- Hambleton, R. K., (1982). *Item response theory: The three-parameter logistic model. CSE Report No. 220. Los Angeles: Centre for the Study of Evaluation.*
- Hambleton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer.
- Hernard, D. H. (2000). Item response theory. In L. Grim & P. Yarnold (Eds.), *Reading and understanding more multivariate statistics (pp. 67-97)*. Washington, D.C: American Psychological Association.
- Hinkle, D.; Wiersma, W. & Jurs, S. (1998). *Applied statistics for the behavioural sciences (4th ed.)* Boston: Houghton Mifflin.
- Norusis, M. J. (1990). *SPSS advanced statistics student guide.* Chicago: SPSS Inc.
- Schumacker, R. E. (2005). Item response theory. *Applied Measurement Associates.*