International Journal of Research in Social Science and Humanities (IJRSS)
Vol 1 No 1 (2020)
www.ijrss.org

One-Parameter Logistic Model and its Application in Test Development

UGADA, Catherine & Dr. EYONG, Emmanuel Ikpi

[1]Department of Educational Foundation, Rivers State University, River State
[2]Department of Educational Foundations and Childhood Education
Cross River University of Technology (CRUTECH)
Calabar’ Nigeria

ABSTRACT

The study focused on the one-parameter logistic model (1PLM) and its’ application in test development. *Two research questions were answered and the design of the study was instrumentation. Instrumentation research is a scientific investigation for meticulous development or construction of a test or measuring instrument that validity measures that concept or psychological construct, which it intends to measure with all accuracy (Kpolovie, 2010). A multi-stage sampling technique was used to acquire a sample size of 200 students for the study. Simple random sampling technique was used to select four schools from the local government area. Stratified random sampling was used to select 50 students each from each of the four schools to give the needed sample of 200 for the study. The instrument for this study was a self-developed mathematics test items. It is a 50 items test. The format is multiple-choice objectives with five (5) options lettered A-E. Findings revealed that the use of one parameter latent trait theory (Rasch model) offers opportunity to deal with core measurement issues such as construct validity as well as providing richer interpretation regarding examinee performance.*

Keywords: logistic model, Psychology, Rasch model, Sampling technique.

INTRODUCTION

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, 2012). IRT assumes that there is a single ability or dominant factor that explains performance, this ability parameter is often denoted as 0. 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 of this study centered on explaining the theoretical perspective of The normal ogive otherwise known as the one parameter logistic model (IPLM), or the item characteristic curve and an illustrative example of the calculation of IPLM will also be shown.

STATEMENT OF PROBLEM

The most reliable means of assessing teaching and learning activities is by administering tests to the students: To maximize testing, one should aim to integrate all the major components of a course content, instruction, objectives, assessment and evaluation. However, psychological traits such as ability or proficiency are constructs. They are unobservable but can be measured directly using a tool called test. The design of tests to measure constructs, however, presents several problems. Since the measurement of psychological constructs is always done indirectly, there is always the possibility that researchers will select different types of behaviour to measure the same construct. As a consequence, different inferences will be concluded. Lack of well defined units in the measurement scale also poses problem. For example, an examinee who is unable to answer any test item does not mean that he or she has zero ability. Instead, all the items have difficulty index which is more than the examinees ability. The study of measurement problems and methods to overcome them is known as test theory. Test theories relate observable traits (such as test score) with unobservable traits (such as ability proficiency) for a measured construct using mathematics model. Therefore, there is the need to consider the normal ogive or the one-parameter logistic model (1plm) and its' application in test development characterization of one parameter logistic (IPL) model.

Purpose of the study

The aim of the study is to determine the one-parameter logistic model (1PLM) and its' application in test development characterization of one parameter logistic (IPL) model. Specifically, the study intends to:

1. Examine the place of one parameter model in testing
3. Assess the application of one parameter logistic (IPL) model in test development

Research questions

The following research questions guided the study:

1. Where is the place of one parameter model in a normal ogive curve
3. How is the normal ogive curve applied in test development?

LITERATURE REVIEW

Theoretical model development of the place of the one parameter model

The item response model was built based on the one parameter model, which is a standardized form of the one parameter model which is particularly useful in determining the various percentile point in a distribution of scores (Hinkle, Wiersma & Jurs, 2008). 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:

According to Crocker and Algina (2016) 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, 2016). 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.

Item characteristic curve (ICC) for a one-parameter model and its 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:

1. Item statistics that are independent of the sample from which they are estimated.
2. Examinee scores are independent of test difficulty.
3. Item analysis accommodates matching test score to examinee knowledge level.
4. Test analysis does not require strict parallel test for accessing reliability.
5. 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, 2013).

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 that the scores from the different forms are comparable, test score equating are usually carried. The 1- parameter model believes that the ability (Ø) 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 baking 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

RESEARCH METHODOLOGY

The design of the study is instrumentation. Instrumentation research is a scientific investigation for meticulous development or construction of a test or measuring instrument that validity measures that concept or psychological construct, which it intends to measure with all accuracy (Kpolovie, 2010). The design is appropriate since the study involves construct validity of mathematics test items. A multi-stage sampling technique was used to acquire a sample size of 200 students for the study. Simple random sampling technique was used to select four schools from the local government area. Stratified random sampling was used to select 50 students each from each of the four schools to give the needed sample of 200 for the study. The instrument for this study was a self-developed mathematics test items. It is a 50 items test. The format is multiple-choice objectives with five (5) options lettered A-E. The item for the study is drawn from the senior secondary school (S.S.S.3) three syllabus. Effort was made to ensure that the topics for which the items were draw were those covered by the students in all schools selected. Test blue print was used to ensure content coverage. Experts in the field of mathematics verified the instrument. These were done to ensure both content and face validity. Some items were deleted while some were reconstructed which led to the emergence of 150 items from the 200 items originally developed. They were administered to 50 student who were not part of the sample used for the trial testing of the instrument. After the administration a graph of the normal curve was the plotted and used in the study.

Method of data analysis

The students’ responses in the final test items from the sampled schools were prepared for the analyses using a Rasch model software, WINSTEPS version 3.75. In WINSTEPS, the measures are determined through iterative calibration of both person and item using the Mathematics Achievement test. which was then used to produce a normal ogive curve.

RESULTS

The result depicts the scores of the sample students in mathematics when converted to produce a typical normal curve as presented below.

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.

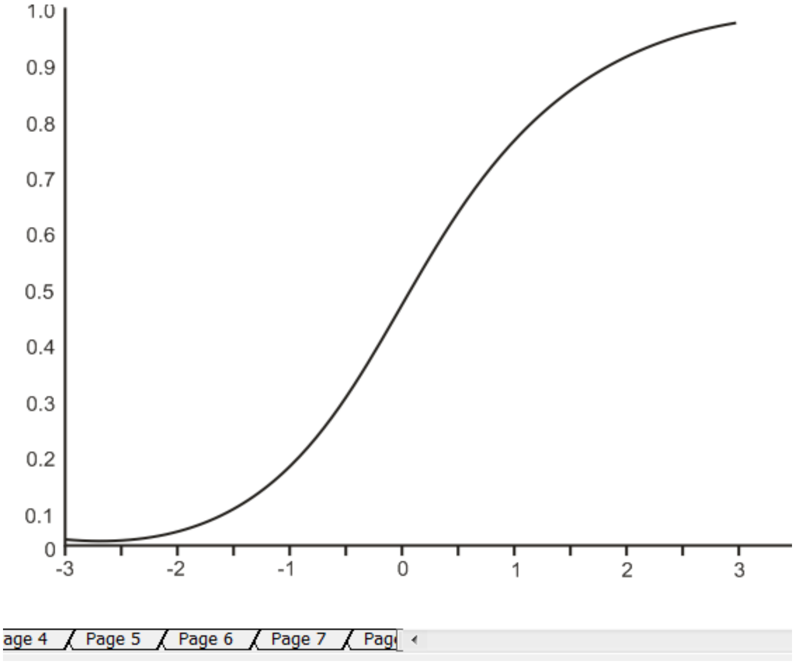

Figure 1: Normal ogive

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 Ø (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). 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 Ø. 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: 1/∅V) = $ Which expresses that the probability of a correct (Yi = 1) response to an item i is a function of ability of the examinee (Ø), the threshold or difficulty parameter bi and the discrimination parameter a. the slope as mentioned earlier is fixed at 1 for all items.

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 ∅, 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 (2015) explained that the characteristics of an item response model (IRM) involve four ideas:

1. An IRT model must specify the relationship between the observed response and the underlying unobservable construct;
2. The model must specify ways of estimating scores on the ability;
3. The underlying unobservable construct can be estimated based on the examinee's scores
4. 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.

ONE PARAMETER MODEL IN TEST DEVELOPMENT

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 Ø. 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(Ø), 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 (PØ) 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 (IRP) 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.

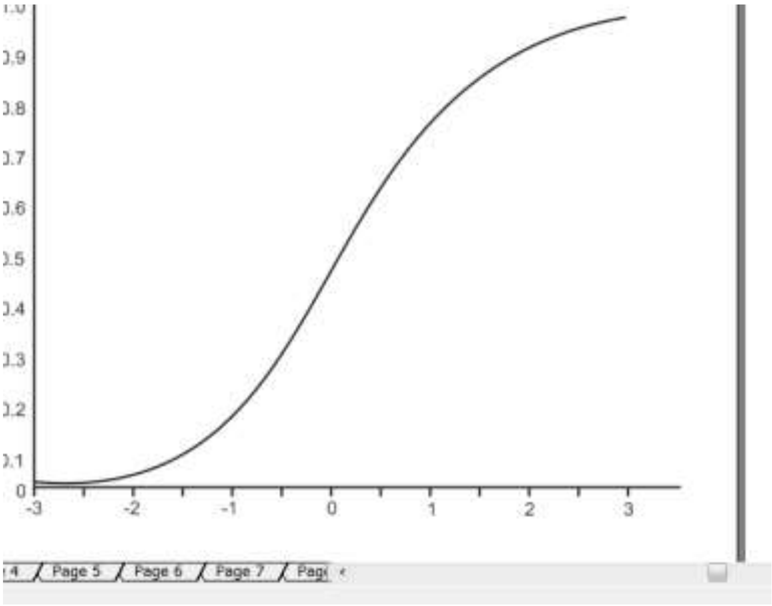

Figure 2: A typical item Characteristic Curve

The ICC shows a near ideal case and indicates that when Ø 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 Ø 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.

DISCUSSION OF FINDINGS

The one parameter latent trait model analyses with the one parameter model as presented in table 1 indicated that both means values were close to the expected value of 1.00. The findings of the study proved that the IRT must be a basic necessity in test development. As the Ø 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 and Algina (1986) explained that two interpretations can be discerned from IRT. Thus, the use of one parameter latent trait theory (Rasch model) offers opportunity to deal with core measurement issues such as construct validity as well as providing richer interpretation regarding examinee performance. Theoretically, this study has added more evidence in favour of the one parameter latent trait theory as having the capacity to resolve some of the rudimentary issues in measurement. However, in order for construct validity to hold, the theory requires more evidence. Test developers would have to have a thorough understanding of the measured construct. This one parameter latent trait model analysis has provided useful information which not only can be used for future developments, modification and monitoring achievement assessments, but also for establishing a process of validating pedagogical assessment.

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. (2016). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Hambleton, R. K. & Jones, R. W. (2013). 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. (2005). *Item response theory: Principles and applications.* Boston: Kluwer.

Hambleton, R. K., (2012). *Item response theory: The three-parameter logistic model. CSE Report No. 220. Los Angeles: Centre for the Study of Evaluation.*

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, Wiersma & Jurs, 2008). *The theory and Practice of item response theory.* New York: The Guildford Press. Advances in Social Sciences Research Journal (ASSRJ) 4, (12), 13-18.

Kpolovie, P.J. (2010). Advanced research methods. New Owerri: Springfield publishers ltd.

Schumacker, R. E. (2005). Item response theory. *Applied Measurement Associates.*