Chapter 3

CONCEPTUAL UNDERSTANDING OF RASCH MODEL

3.1 Introduction

3.2 Latent Trait Theory / Item Response Theory

3.3 Rasch Measurement Models

3.4 Importance of Item Analysis

3.5 Basic Theories of Item Analysis
   3.5.1 Classical Test Theory
      3.5.1.1 Test Construction and Analysis
              According to Classical Theory
   3.5.2 Latent Trait Measurement Model

3.6 Features of Latent Trait Models
   3.6.1 Item Characteristic Curve (ICC)
   3.6.2 Dimensionality of the Latent Space
   3.6.3 Information Function
   3.6.4 Local Independence
   3.6.5 Estimation Procedure
   3.6.6 Test Construction

3.7 Application of Rasch Measurement
   3.7.1 Item Banking
   3.7.2 Test Design
   3.7.3 Self Tailoring
3.7.4 Response Validation
3.7.5 Item Bias
3.7.6 Individual Diagnosis
3.8 Further Innovations in Rasch Model
3.9 Resume
3.1 INTRODUCTION

Many psychometric tests have been developed in India and elsewhere by using the classical test theories in which two parameters of the test items have been considered. BAS is the single parameter test which is based on Rasch Model. The Rasch Model is a specific instance of the Latent Trait theory or Item Response theory (LTT/IRT). So it is pertinent for the investigator to discuss LTT/IRT in full detail. The subsequent paragraphs would deal with the different aspects of LTT with specific reference to Rasch model as practised in BAS.

3.2 LATENT TRAIT THEORY / ITEM RESPONSE THEORY

Item response is related to a trait which is assumed to be an ability of the person. This relation between ability and the item response could be described by a normal ogive model. The implications of this ogive model were difficult to understand up to 1930s but present day item response theory has been developed mainly on the basis of the logistic function, which gives results closely similar to those from the normal ogive model. IRT admits several alternative models based on the logistic function, differing mainly in the number of parameters that are required to describe the function.
The one parameter model is called the Rasch model because the logistic function of the normal ogive model fully satisfies the one parameter. This model is also called the unidimensional model because of only one parameter used in explaining the normal ogive model.

There are other models also but they do not come under Rasch model because they are not based on single parameter of difficulty. This model which are explained by three parameters of position, shape and guessing is called Birnbaum model of multidimensionality. The present investigator was not concerned with the Birnbaum model. Hence she would only discuss the Rasch Model fully in subsequent paragraphs that follow.

3.3 RASCH MEASUREMENT MODELS

The "Rasch Measurement" models developed by Danish Mathematician George Rasch between 1951 and 1959 and explained in his 1960 book, "Probabilistic Models for some Intelligence and Attainment Tests" are the most important advance in psychometrics since Thurstone.

Objective measurement depends on measuring tools which function independently of the objects measured. This requires a response model for calibrating their...
functioning which can separate the effects of tool and object. Rasch was the first psychometrician to realize the necessity and sufficiency for objectivity of logistic response models with no interaction terms. They exemplify the principles of measurement on which all scientific objectivity is based.

Rasch models are practical realizations of "fundamental measurement". When data can be selected and organized to fit a Rasch model, the cancellation axiom of additive conjoint measurement is satisfied, a perfect Guttman order of response probabilities and hence of item and person parameter is established, and items are calibrated and persons measurement on common interval scale.

There are several models of "Rasch measurement models".

Rasch models:

There are four models under this heading, according to Wright B.D.¹

(1) Rating Scale Model
(2) Poisson Model
(3) Partial Credit Model
(4) Item Analysis Model

: 61 :
BAS uses Item Analysis Model for estimation of difficulty values of the items and ability calibration of the persons.

When there are only two alternatives so that \( m = 1 \), the model is the simple logistic "Rasch model" so widely used for the sample-free calibration of education test items and the test-free measurement of individual attainments.²

3.4 IMPORTANCE OF ITEM ANALYSIS

The construction of a measuring tool is a complex process which involves many techniques. The writing of items is an important component of the developmental process of creating a measuring test. For doing so the test writer must have a complete grasp of the content as well as different forms of the questions he/she wants to utilize in the test.

The classical source on the techniques of item writing was developed and documented in "Achievement Testing" by Hawkes, Lindquist, and Mann³ and further corroborated by Wesman⁴. Even if these techniques are worked upon, an item may be an enigma. It is very difficult to evaluate an item as "bad" or "good" by just seeing it. The content
of the item may be alright, the distractors may be quite appropriate and the words used in making of the items may be at the appropriate level of the target population, yet the item may not be good in some sense. It is in this context that item analysis have been developed to provide necessary information which can tell whether item is good or not.

3.5 BASIC THEORIES OF ITEM ANALYSIS

There are two basic theories of an item properties or parameters.

(1) Item difficulty, a measure of whether the item is easy or hard for the pupils; and

(2) Item discrimination, the degree to which an item can be distinguished between pupils having different criterion variable scores.

Thus an item can be described as easy (of low difficulty) or hard (of high difficulty) and as one that has a particular degree of discriminating power.

In the test development process the test is administered to a random sample from the target population and the scores are analyzed by using T.L. Kelly's 27% process.
By further analysis scores will yield the two parameters of difficulty and discrimination of the item.

It should be noted here that the numerical values of item difficulty and item discrimination do not intrinsically define a good or bad item. Rather they represent two technical pieces of information that must be evaluated within the overall context of the item.

There are two different measurement theories upon which the item parameters are based viz. classical test theory and latent trait / item response theory. Because each of these theories leads to a unique definition of the item parameters and their interpretations, separate discussion is necessary and it is given in the subsequent paragraphs which now follow.

3.5.1 Classical Test Theory

The basic item analysis model under classical test theory is a correlational one in which an item analysis variable is correlated with a criterion variable. The item variable is a hypothetical variable that is conceptualized as a wrong to right continuum. The criterion variable is the trait being measured by the test. In practice, however, the criterion variable is usually the total
test score and in some applications it can be a variable external to the test such as grade-point average. Assuming there is a bivariate normal relationship between the item and criterion variable, the two item parameters are defined as follows:

(1) item difficulty, the proportion of the population of examinees who answer the item correctly; and

(2) item discrimination, the product-moment correlation between the item and criterion variables for the population of examinees responding the item.

In actual practice these parameters are estimated by sample statistics and an estimate of item difficulty is simply the proportion of the group of examinees who answered the item correctly. A value of .10 corresponds to a difficult item whereas a value of .90 would be associated with an easy item. Thus the index is more properly a measure of item easiness, than of item difficulty. However, tradition dictates that the latter terminology be used.

Because the item discrimination index involves the hypothetical item variable that cannot be measured directly, correlations other than the product-moment are used. The most popular, biserial r is a function of the assumption
made as to how the observed dichotomous item response is related to the hypothetical item variable.

With biserial \( r \), the assumption made is that the hypothetical item variable is normally distributed but dichotomized by a value, \( \gamma \), that delimits an area equal to the item difficulty. In reality the only data available are on whether or not an item was answered correctly. Thus the correlation is between the criterion variable and a dichotomous \((1,0)\) variable.

The numerical values of biserial \( r \) ranges from -1 to +1, but if underlying assumption are violated, values greater than unity are possible.

The user of item analysis techniques must keep in mind that the obtained item statistics are estimators of the actual parameter values. As such they will have sampling distributions and thus a given value may not be replicated in subsequent testing. Baker (1965) has studied the sampling distribution of several item discrimination indices and has shown that they have considerable variability when small samples \((N = 60)\) are used, and thus values obtained on the bases of these small samples should be interpreted with caution.\(^5\)
An important part of test construction process is the establishment of a pool of items for which item statistics are available. Such a pool of 'precalibrated items' enables the user to select items from the pool to meet the specifications in the underlying test plan. The statistics of the items selected are used to obtain a desired test score distribution, assuming the population of examinees is similar to that used to precalibrate the items. For example, the distribution of the values of the item difficulties has a direct impact on the test score distribution. Ebel (1965) has shown that clustering the item difficulties, around a value of 0.5 yields a symmetric test score distribution with a moderate amount of spread. When the item difficulties are uniformly distributed over the range .10 to .90, the test score distribution has a smaller variance than in the previous case. The use of only extreme values of item difficulties (.10 and .90) results in a test score distribution that is very peaked near its mean and has a smaller variance than either of the previous two cases.

From these results it is clear that the wider the distribution of the item difficulty values in a test, the
smaller the variance of the test scores and the lower the value of the reliability coefficient. 7

The spread in value of the item discrimination indices is not so important as their average level. The higher the level, the greater the discrimination among the examinees. Since internal consistency reliability depends upon the separability of examinees, the higher the average value of the discrimination indices in a test, the higher the internal consistency reliability coefficient and the larger the test score variance.

When precalibrated items are available, the test constructor has considerable flexibility in selecting items to achieve desired test characteristics. However, the test statistics computed from the item statistics are simply preliminary indicators. Only when the test has been administered and analyzed one can obtain the actual test statistics. But when proper test development procedures are employed the predicted and obtained test statistics should show reasonable agreement. Such agreement is hardly to be found in tests developed under classical theory or true-score model, because in the true-score model the units of X (and of true-score T) do not lie on an interval scale, under such circumstances, the increase in ability arising from one more item correct on test A is

: 68 :
not necessarily the same as that from one more item correct on test B (even when A and B are designed to measure the same ability and have the same number of items). The very use of word "ability" presupposes the existence of an underlying scale (a latent trait), with some equal interval units.

From the above discussion, it is seen that the true-score or classical theory has been considered too restrictive, unsubstantiated or of inadequate scope for the measurement required in many educational and psychological circumstances. 8

3.5.2 Latent Trait Measurement Model

In the classical test theory, the difficulty index of the item is not anchored to the ability of the examinee. This is the greatest flaw of the classical test theory.

Since the publication of the fourth edition of the Encyclopedia of Educational Research (1969), a modern theory of measurement called Latent Trait or Item Response Theory (LTT/IRT) has come into wide use. Actually this theory has been evolving since the turn of the century, but its recent emergence as a significant element in measurement practice can be attributed primarily to two factors:
(1) Its conceptual superiority over the classical approach, and
(2) the availability of computer programmes that implement the parameter estimation procedures.

Since classical test theory is a theory of test scores, item analysis is a set of techniques that is rather disjoint from the basic theory. In sharp contrast LTT/IRT is based on the parameters of the items that make up the test. As a result the item analysis techniques are inextricably interwined with the theory itself, and so it is necessary to have a basic understanding of the theory in order to use these techniques. The basic concepts of the theory underlying the use of item analysis are described in the subsequent paragraphs.

3.6 FEATURES OF LATENT TRAIT MODELS

According to Hambleton and others (1978), there are three important notions in the general theory of latent traits which form the basis of the assumptions of these models. The specific aspects of LTT which apply to Rasch Model will only be emphasized in the discussion.

A latent trait is a psychological dimension (construct) whose existence is postulated to account for replicable
variation in observation of behaviour in certain well-articulated situations. There is no question of physical existence for psychological traits; the value of any trait lies wholly in its utility as a tool for understanding experience through behaviour. Since a trait is an abstraction; it may be manipulated, restricted, varied and otherwise transformed according to whatever purpose is to be served.

Latent trait models are probabilistic models which describe the probability relationship between observations and unknown parameters representing the traits. The probabilistic aspect insures that the mathematical form of the model provides just the probability and not the certainty of occurrence of prescribed events.

3.6.1 Item Characteristic Curve (ICC)

Latent trait models which go under the name of 'Rasch' models are derived by the Danish mathematician, George Rasch, and have been promoted in the U.S.A. by B.D. Wright in Chicago and in Europe by G. Fischer in Vienna.

For binary scored data, the Rasch model has the form:

\[
P (X_{vi} = 1; \theta_v, d_i) = \frac{\text{Exp.} \left( \theta_v - d_i \right)}{1 + \exp \left( \theta_v - d_i \right)}
\]
Rasch arrived at the shape of the item characteristic curve from the above equation.

The basic building block of LTT/IRT is an item, and the technical characteristics of an item are couched in terms of parameters of an item characteristic curve.

This curve is the functional relation between the probability of correct response to an item and a hypothetical latent trait. The latter denoted by $\theta$ and referred to generically as "ability" is a construct such as intelligence, scholastic ability, or arithmetic ability that the item is attempting to measure. The functional relationship is best depicted graphically as in fig. 3.1.
Fig. (3.1) Item Characteristic Curve, $B = 0.5$, $\alpha = 1.0$
The horizontal axis is the ability scale. It has a midpoint at zero, a unit of measurement equal to one, and a theoretical range from \(-\infty\) to \(+\infty\).

The vertical axis is the probability of correct response to the item, \(P_i(\theta_j)\), where \(i\) indexes the item and \(j\) the ability level. The curved line relating the two variables is the item characteristic curve.

The basic ICC model is the normal ogive model. Models for the ICC other than the normal ogive are also employed in practice. The logistic ogive in which the abilities of persons are in logarithmic units is widely used and is defined by

\[
P_i(\theta_j) = \frac{1}{1 + \exp \left( -a_i (\theta_j - B_i) \right)}
\]

where \(a_i\) is the point on the ability scale where the probability of correct response is one-half for examinees of that ability.

\(B_i\) is referred to as the item difficulty parameter. This difficulty parameter has theoretical limits of \(-\infty\) to \(+\infty\) but obtained values fall between \(-4\) to \(+4\). A hard item will have a value of \(B_i > 0\) and an easy item a value of \(B_i < 0\).
The item discrimination index $a_i$ is defined as the reciprocal of the standard deviation of the normal ogive. A rather flat ogive has an underlying $\sigma$ that is numerically large and a steep ogive has a small $\sigma$ associated with it, thus an item that discriminates sharply among ability levels will yield a large numerical of $a_i$. Although the theoretical range of $a_i$ is from $-\infty$ to $+\infty$, the value is usually in the range $0 \leq a_i \leq 2.5$ and negative values of $a_i$ are undesirable for the correct response to an item. For the interpretation purpose $a_i$ can be thought of as the slope of the ICC at the point $B_i$.

Given an ICC model and these two parameters, the relationship between ability level and probability of correct response to an item can be completely specified. Despite the use of common terminology the item difficulty parameter of LTT/IRT has a completely different meaning from that of classical test theory. It is not a measure of the proportion of examinees getting the item correct.

The logistic model of ICC has the advantage over the normal ogive of not involving an integral. If the value of the logistic discrimination parameter $a_i$ is multiplied by 1.702, the logistic ogive specified will match the normal ogive having the same value of $a_i$ to within .01 over the full ability scale. As a result the logistic model is
often substituted for the normal ogive to reduce the computational burden of item analysis. This model has been used by Rasch for unidimensional items.

3.6.2 Dimensionality of the Latent Space

This notion refers to the number of latent traits that underlie the performance of persons undertaking the test. Most latent trait models assume that single latent ability is sufficient to explain or account for individual differences in performance. Models that make this assumption are referred to as "unidimensional" models. Lord (1968) states that this assumption concerning the unidimensional nature of a set of items is not strictly true for most tests, although tests in a number of cases may reasonably approximate to this assumption. Various authors, such as Lord (1968) and Lumsden (1961), suggest that factor analysis provide a useful method of assessing the unidimensionality of a set of test items. Fortunately, a better statistical method of assessing unidimensionality has been developed namely, goodness-of-fit to the Rasch Model. Lumsden (1978) makes the point that if the unidimensionality requirement is met, the Rasch model will be realized.
For items to be unidimensional, it would be necessary that a single item parameter (item difficulty) should account for differences in the probability of persons getting items right. This is so in the case of the Rasch model, which can therefore be characterized a fully unidimensional model.

In the case of other latent trait models, it has been found necessary to introduce additional parameters. In a two-parameter latent trait model, the second parameter in addition to the item difficulty is item discrimination. This is introduced to take account of the fact that in some tests certain items will discriminate better between high and low scoring persons than others.

In the case of three-parameters latent trait models, in addition to the item difficulty and item discrimination parameters, a third parameter usually for guessing is introduced. Guessing parameters are sometimes required to take account of individuals guessing the correct answers to items, as in multiple-choice tests.

Such tests as are containing more than one parameters can not be called unidimensional.

In view of these considerations, the term "unidimensional" will be used in the full sense of both items and
persons being unidimensional. To repeat, the Rasch model is the only unidimensional latent trait model in this sense.

In constructing fully unidimensional tests, Levy's (1973) comment that what is needed are more tests related to theories rather than theories related to tests, is particularly relevant.¹⁶

To demonstrate unidimensionality, it is necessary both

(1) to specify the item types so as to ensure, from the point of view of an educational or psychological examination of the items, that they form a unitary, unidimensional set, and

(2) to ensure that the items in the set fulfil a statistical criterion of unidimensionality namely, goodness-of-fit to the Rasch model.

In terms of contents the test constructor should be theoretically satisfied that all the items are homogeneous and measuring the same thing; i.e. that they require the same combination of processes from the testee for their response, and that the responses to a single item or to different items should not be qualitatively different.
A good example of a unidimensional test is one of word reading in BAS.\textsuperscript{17} In such a test, each item consists of a printed word, which a child looks at and to which one oral response is made. It is clear that this is a complex task because it involves many cognitive processes, including visual perception, search and retrieval from long-term verbal memory stores and verbal encoding of the responses. All of these processes are in themselves, complex, and yet the test is unidimensional in that each item broadly requires the operation of the same set of responses.

The test constructor ensures that these are as uniform as possible by deleting items which sample some other process or area of knowledge. Having ensured that the items form a uniform set, the test constructor then needs to be satisfied on the statistical criterion. Both these criteria - the Psychological/Educational and Statistical - are necessary conditions for unidimensionality, but neither is sufficient in itself.

The criterion of goodness-of-fit is concerned with establishing that the items in a theoretically unidimensional test keep their relative difficulty values across all ability groups of individual for whom they are intended.\textsuperscript{18} That is why it is said that Rasch model is a necessary but not sufficient condition for unidimensionality.
3.6.3 Information Function

In latent trait model, ability of a person is linked with difficulty index of the test item. So one can conceptualize a conditional distribution of ability estimates, \( \theta \) yielded by examinees who share a common underlying ability level \( \theta \). This conditional distribution will have a mean \( \theta \) and variance \( \sigma^2_\theta \). This amount of information is a concept due to Sir R.A. Fisher and is defined as \( I(\theta) = 1/\sigma^2_\theta \). Thus the larger the variance of the \( \theta \) at a given \( \theta \) level, the less information one has about an examinee's unknown ability level. Under LTT/IRT the amount of information can be obtained for a test and for the individual item in a test at each ability level.

The amount of information contributed by an individual item is given by the following formula under a logistic ICC model:

\[
I(\theta_j) = \sum_{i=1}^{n} \frac{2}{a} P_i(\theta_j) Q_i(\theta_j)
\]

where \( P_i(\theta_j) \) is obtained by evaluating the item characteristic curve model at \( \theta_j \).

The test information function plays a role in LTT/IRT that is related to that of the reliability coefficient in
classical test theory. The usual reliability coefficient is defined as the product-moment correlation between test scores on two parallel tests. As such it depends not only on the test but also on the group of subjects taking the test.

Under LTT/IRT the situation is radically different in that the test information function is intrinsic to the test, depending only upon the ability scale and the items in the test.

Since the test information function is evaluated at each ability level, the amount of information will be unique to each level. Thus under LTT/IRT one uses the test information function to examine how well the test is estimating ability over the range of interest. Consequently the information function is much more powerful concept than reliability, both theoretically and in practice.

3.6.4 Local Independence

The assumption of local independence is that the probability of a person of a given ability level answering a test item correctly is not affected by him or her performance on any other item in the test. Since human beings are changeable and learn as a result of experience and
practice, this assumption is unlikely to be strictly satisfied in many educational or psychological tests. Test constructors attempt to deal with this difficulty sensibly, however, by providing an adequate number of practice items at the start of a test and by ensuring that items in a test are arranged in ascending order of difficulty, with appropriate additional instructions inserted in test at any point if new complexity has been added to the items. As a result test items are seldom administered randomly.

The chief test application of this assumption may be of crucial importance in the development of item banks. These involve the calibration of a relatively large number of items, a selection of which can be drawn from the bank and assembled together to form a test for a particular purpose.

If the difficulty values of items change when they are administered within different test item sets to different samples, this may be because the items are not locally independent. It is possible to test the goodness-of-fit to the Rasch model of a set of items which is common to two different tests, these common items are called link-items, since they enable the different tests to be linked together onto a common scale provided that they fit the Rasch model.
3.6.5 Estimation Procedure

Under LTT/IRT two sets of parameters must be estimated from the item response data. Assuming a two parameter model such as the normal or logistic for the ICC, there will be two n-item parameters to be estimated for a test having n items. Since classical test theory was a theory of test scores, the only information desired about an examinee was his raw test score. However, under LTT/IRT the raw test score has little intrinsic value as the goal is to estimate an examinee's ability directly. Thus the second set of parameters to be estimated are the ability levels of the N examinees who respond to the test items, and so a total of two n + N parameters are to be estimated.

The estimation of the necessary parameters is a computationally complex process that can only be accomplished by digital computers.

3.6.6 Test Construction

All of the key elements for the application of Rasch model on the bases of dichotomously scored items would be made available.

The data yielded by an administration of a test can be analyzed and interpreted in an adapted test that meets
"a priori" specifications can be constructed from pools of precalibrated items, and new and unique testing procedures can be developed.

With LTT/IRT the test norming process is known as test calibration. To calibrate a test a population of subjects is used as a calibration group. The item response data obtained by administering the test to the calibration group are analyzed via the estimation procedure. The results are the estimated values of the item parameters $a_i$, $B_i$ for each of the $n$ items and the estimated ability levels $\theta_i$ for the $N$ examinees expressed in a matrix that has a midpoint of zero and a unit of measurement of one for that group. Since the basis of these estimates is a calibration group, they are treated as if they were the parameters. In the terminology of LTT/IRT the items have been precalibrated, that is, their parameters are known.

One of the most powerful features of LTT/IRT is the conceptual framework it provides for the test construction process. When an item pool containing precalibrated items is available, LTT/IRT furnishes the test constructor with a versatile repertoire of techniques based on the ICC parameters $a_i$, $B_i$ and the test information function.
These two facts of the theory allow the test constructor to devise tests for very specialized as well as very broad purposes.

3.7 APPLICATIONS OF RASCH MEASUREMENT

The applications of Rasch model are legion, out of which the following seem to be the most important.

3.7.1 Item Banking

When a large number of test items are constructed so that they calibrate along a single dimension, and when they are used so, they retain these calibrations over a useful realm of application, then a scientific tool of great simplicity and far-reaching potential becomes available. Its item contents serve the composition of an infinite variety of pre-equated tests. Neither the difficulty nor shape of these tests need have any effect on their equating. All possible scores on all possible tests are automatically equated in the measures they imply through the common calibrations of their bank items. Whatever the test, its measures are expressed on the common variable defined by the bank. Furthermore the validity of these calibrations and of each measure made with bank items can be verified at every step.
3.7.2 Test Design

The positioning of items along the dimension they define makes test design easy. Tests can be targeted on any region along the variable represented by calibrated items. The item chosen for a particular test can be spread over the target region in whatever way is most informative. The best designs are obtained by bunching items at decision points to maximize decision information and by spreading them evenly over targets to maximize decision information.

3.7.3 Self-Tailoring

Persons can also make their own choice of item difficulty as they go along. The items in their test can be arranged to increase in difficulty. People can choose their own starting point. If they feel strong, they may work ahead into harder items until they reach their limit. If they feel weak, they may start with easy items.

3.7.4 Response Validation

The analysis of fit enables the validity of each response to be examined. This is an important step in estimating a measure from test performance. The items
used will vary in their positions along the variable. This will happen when items are spread to cover the target. It will also be forced by limitations in item resources.

When items vary in their difficulty, persons are expected to do better on easier items than on harder ones. Because the response model is explicit in this regard, this expectation can be formulated into an analysis of fit for any response pattern. This enables the validity of each and every test performance to be examined before any measures estimated from it are reported.

3.7.5 Item Bias

The analysis of response pattern fit allows each person's item responses to be diagnosed in detail. If any theory is processed that classifies them by response format, page layout, booklet location, item text, topic, or approach, then it is possible to calculate how much each person's responses are disturbed by these categories.

When a disturbance is found, it is possible to estimate the extent to which the unusual category is biased for each person. There is no other objective basis for the analysis of item or test bias. Bias estimated from groups can never satisfy the right of each individual to be fairly treated regardless of membership.
3.7.6 Individual Diagnosis

More important is the identification of each test taker's strengths and weaknesses and the use of this diagnosis to find what he or she needs next. Most test takers are associated with programmers dedicated to improving them. The justification for testing is the intention to use tests to help test takers. For this, an item content diagnosis of each test taker's response pattern is essential. Since the response residuals from the measurement model manifest all the diagnostic information the test contains, their analysis is also all that can be done statistically.

3.8 FURTHER INNOVATIONS IN RASCH MODEL

Upto this point, it was assumed that an item was dichotomously scored a choice that is somewhat arbitrary. There are many testing situations in which one could properly use a 'graded' response. In such cases the hypothetical item response variable is divided into three or more ordered categories and an examinee's response is assigned to a response category. For example, one could categorize the possible responses with regard to their level of understanding of a concept. Because the response categories are mutually exclusive and exhaustive,
response category probabilities at any point on the ability scale must sum to unity.

In addition, the requirement that the response categories be ordered means that the item difficulty parameters must also be ordered. "Samejima has published extensively on the theoretical aspects of the graded response and has used both normal and logistic models for the ICCs". 20

The availability of item estimation techniques for the graded response case makes item analysis of rating scales possible. Since most scales involve a single underlying variable, the examinees can be located on a latent trait continuum and the item response categories can also be located along the same continuum. If the item functions properly, the difficulty parameters of the item response categories should be neatly ordered along that continuum. Unevenly spaced or improperly sequenced values of difficulty parameters would identify less than optimal items.

A group of European psychometricians has been developing an extension of the Rasch model known as the linear logistic test model (LLTM). The starting point of this model is the usual one-parameter logistic model.
Each item in a test is considered to entail a common set of cognitive operations. For example, the difficulty of elementary school arithmetic problems may be a function of the student's recognition of the type of problem, the magnitude of the numbers involved, the complexity of the procedure, etc. Each of these factors is considered a cognitive operation and has a parameter associated with it.

3.9 RESUME

The item analysis technique of classical theory have been used for more than half century. These procedures are well known and simplified explanations are given in statistical and mental measurement books.

The most serious defect of the classical item parameters is that they are group dependent. Thus the indices can be interpreted only within the context of a given group examinees. Despite these limitations, the experienced measurement person develops a clinical skill in interpreting the values of the item statistics in many different context.

As a relative newcomer to the measurement scenario, the basic precepts of the latent trait theory is not widely known nor its applications widespread among practitioners. Nonetheless, LTT/IRT is the state of the art in measurement
Because it is based on the parameters of the test items, item analysis plays a prominent role. The item parameters, $B_i$, $a_i$ have simple yet powerful interpretations that greatly facilitate the understanding of test results.

In addition, the item parameters can be invariant with respect to the group taking the test.

LTT/IRT is also capable of extensions such as to graded and nominal response models as well as to multi-component models. As a result of this richness, LTT/IRT should be the standard basis for item analysis.

Thus, BAS which is based on LTT model and particularized on Rasch model is a scientific tool which has immense reliability and validity. This is the reason why the present researcher has been tempted for its adaptation for the children of Gujarat.
REFERENCES


7. Ibid., p.960.


18. Ibid., p.62.
