CHAPTER - I

Introduction
# CHAPTER I – INTRODUCTION

<table>
<thead>
<tr>
<th>Detail</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Technology in Education</td>
<td>2</td>
</tr>
<tr>
<td>1.3. Computer Based Testing (CBT)</td>
<td>2</td>
</tr>
<tr>
<td>1.4. Need for the Study</td>
<td>4</td>
</tr>
<tr>
<td>1.5. Problem Explanation</td>
<td>6</td>
</tr>
<tr>
<td>1.6. Significance of the Study</td>
<td>7</td>
</tr>
<tr>
<td>1.7. Computerized Testing</td>
<td>11</td>
</tr>
<tr>
<td>1.7.1. Technology in Monitoring</td>
<td>11</td>
</tr>
<tr>
<td>1.7.2. Models of Adaptive Test</td>
<td>12</td>
</tr>
<tr>
<td>1.7.3. Computer Adaptive Test (CAT)</td>
<td>13</td>
</tr>
<tr>
<td>1.7.4. Estimation of Ability</td>
<td>15</td>
</tr>
<tr>
<td>1.7.5. Item Response Theory (IRT)</td>
<td>17</td>
</tr>
<tr>
<td>1.7.6. Application of Item Response Theory</td>
<td>18</td>
</tr>
<tr>
<td>1.7.7. Performance Mapping</td>
<td>19</td>
</tr>
<tr>
<td>1.8. CAT Package Development</td>
<td>20</td>
</tr>
<tr>
<td>1.8.1. The Computer as a Means</td>
<td>22</td>
</tr>
<tr>
<td>1.8.2. Potentials of the Present CAT Package</td>
<td>23</td>
</tr>
<tr>
<td>1.9. Variables of the Study</td>
<td>24</td>
</tr>
<tr>
<td>1.9.1. Dependent Variables</td>
<td>25</td>
</tr>
<tr>
<td>1.9.2. Independent Variables</td>
<td>26</td>
</tr>
<tr>
<td>1.10. Objectives of the Study</td>
<td>30</td>
</tr>
<tr>
<td>1.11. Organization of the Thesis</td>
<td>31</td>
</tr>
<tr>
<td>1.12. Delimitation of the Study</td>
<td>32</td>
</tr>
<tr>
<td>1.13. Scope of the Study</td>
<td>33</td>
</tr>
<tr>
<td>1.14. Conclusion</td>
<td>33</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

1.1. INTRODUCTION

To develop more effective learning skill and achievement; government has been embarking on substantial programs of reforms. For developing countries the educational requirement is more, because of over population and mass education. The teaching and learning processes have been dramatically altered by the convergence of varieties of technological, instructional, and pedagogical developments in recent times after the advent of computer (Bonk & King, 1998; Marina, 2001; Smith, 2002; Preety chawla, 2010). What is needed in solving today’s and tomorrow’s educational problem regarding the quality is nothing less than a ‘quantum leap’ a revolutionary change in educational evaluation system. Computer-based assessment remains a fertile field for enrichment of research and development in science (Huang, Chenn-Jung, et.al., 2012).

By developing suitable computer based testing tools, the evaluation will be objective-based and can be able to identify potentially strong and knowledgeable persons. The necessity to use computer in education is resulted from reasons such as, difficulty in education system, rapid increase in the number of students, increase in amount of information, more complicated content and insufficient teachers for the huge mass of learners.

If innovations are introduced in the present system of examination the quality can be improved which in turn improves the education as a whole. Considering the developments in contemporary approaches to computer-based science assessment, it is evident that computers are viable means for identifying the various kinds of factors operating in assessment. Continued research and development efforts are needed in order to use computers as tools that are unquestionably effective to trace any behavioural pattern in assessment of science education. Developing countries like India...
must increasingly look more to the application of computer technology in the evaluation system (Slykhuis David, & Park, John, 2006).

1.2. TECHNOLOGY IN EDUCATION

Technology is used in education for office automation, correspondence, publication of results, admission and minimally for instructional purposes also. From the application point of view technology is relatively minimally used in science evaluation than for science education. Technology in education refers to utilizing technology for teaching-learning and evaluation by directly involving teacher-learner for reaching higher goals of education through enhancing potentials and empowerment process. Researches in technology reveal only the need for quality improvement. The present study is only an epitome in the iceberg of the problem of technology in evaluation in terms of ability estimation and potential identification in certain internal and external variables that are believed to operate with and help in equipping the testees in physics learning and testing.

Following the concepts gathered from technological applications in various areas of education, the focus is turned to Computer Adaptive Test (CAT) for its unique status in assessment. The use of computers as information processing tools, the application of computer technology in user-friendly interactive learning environments, and the possibility of designing instructional tools to meet individual needs of students, will make the computers potentially a powerful tools for assessment, suggests David D. Kumar (1994), Chatzopoulou, & Economides, (2010), Fehr, Charles Norman (2011) and Yen, Yung-Chin., et.al. (2012).

1.3. COMPUTER BASED TESTING (CBT)

Assessment can be the basic for improvement in learning and has become a hot topic among educators across the country. Examinations do not really reflect the true
learning. It is not the benchmark for judging the quality. Several initiatives have contributed to the increased need for quality assessment, which gives information regarding strengths and weaknesses, skills and abilities to plan activities accordingly. The success in the great enterprise of national reconstruction, whose principal objective is to raise the standard of living of our people, depends on the quality of the individuals (Annie Brown, 2012).

Assessment systems must be carefully designed to ensure that they provide information that fulfills the purpose for assessment and to increase the likelihood of improving the quality (Rajesh Hassija, 2012). Traditional test fails to assess the innate potential of the learners. Even technology has limits; it does not have the advantage of increased understanding of specific human abilities. Efforts are made to modernize the traditional test. There has been a significant movement towards measuring specific abilities as alternatives to traditional tests. This new breed of tests includes the computer based testing. Test mode CAI is the first proposed format when computer was employed in educational evaluation and was followed by its application in instruction. The computer based teaching model (Lawrence Stolurow and Daniel Davis, 1965) has a pre tutorial phase which selects a teaching programme for a particular student and a tutorial phase that provides instruction and monitors the student’s performance to discover whether the programme may be more suitable than the previous one.

The aspects of continuous assessment can be automated since automated assessment strategies can assist in estimating abilities of the learner and providing feedbacks (He, Wei, & Reckase, 2008; Holden, 2012). Self-evaluation is generally automated and forms part of formative evaluation that provides instant feedback to the learners. Quiz test is another method of formative evaluation where the learners can
read a question, choose an answer and have the answer locked in. A computer with its capacity for storing huge amounts of text and graphics, its quick access to memory, its video monitor for displaying the items, its keyboard for recording the responses and its enormous power to perform calculations according to the complex rules, forms an ideal platform for adaptive testing (Lina Ashar, 2012).

1.4. NEED FOR THE STUDY

The present system of education in India is too much dominated by examination which dictates what to teach and how to teach. Due to the various defects of the present system of examination, the goal of education is not achieved till now (Rajesh Hassija, 2012). Evaluation should measure the developmental changes among learners. The domination of examination over education was pointed out by several national commissions of education. The Kothari commission report published in 1966 has pointed out the necessity for the improvement of examination system. But in spite of the repeated requests of all these commissions for improving education by reforming examination much has to yet to be done. Computer technology can open up opportunities for developing innovative assessment tools in almost all the fields of education, especially in science education (Lina Ashar, 2012). The nature of computers as information processing tools, user-friendly, interactive learning environments and the possibility of designing instructional tools to meet individual needs of students, make computers potentially powerful tools for assessment.

Any evidence that reveals progress towards the goals of education should be used in evaluation. For the improvement in the quality, the evaluation system has to be improved, by which a thorough mastery of subject matter can be achieved. One such system of evaluation can be achieved by technology in education i.e. Computer Based Testing (CBT). Computer based testing has finally become a viable alternative to
conventional paper and pencil testing. In CBT, the evaluated behavior is mapped and
the progress is tackled and maintain as a cumulative record of progress individually.
CBT is a system of evaluation using computer, which is useful for ability estimation,
potential identification and for quality enhancement.

The performance assessment through feedback mechanism generates changes
among various components of the teaching learning process, thereby leading to
improvement of the learner, the teacher, the teaching learning strategies and even the
administration. In such a way, evaluation should be considered as the service
component of the teaching learning process and used to emphasis regular feedback to
improve teaching and learning. Thus, evaluation is an integral part of the teaching
learning process. It is a continuous, comprehensive, cooperative, mastery oriented,
individual based, activity oriented and objective centered process.

Millennial generation learners expect entertainment in all educational activities,
such as Teaching, learning and Evaluation. They want fair and consistent judgment
from teachers, and they want a clear indication of what they need to do, to be
successful. They also need clarity in their assessment tasks and want clear
unambiguous instrumentations, opportunities to discuss problems with peers and
teachers, and timely feedback.

Improving and sustaining quality in education cannot be achieved without
explicitly focusing the educational systems and processes. Assessment experience may
be stressful. Students are typically anxious about are examination. The only way to
prepare students for the future is by creating an atmosphere, where they want to do the
best for themselves, where they are guided by their natural instincts to excel and
achieve, without much stress and pressure. CBT, which is individualized in nature, can
assess students without having undergone unnecessary stress. Having all the
expectations and standards clearly articulated in CBT, makes the assessment more amenable to Moderation.

As a teacher, constantly interacting with the students and experienced school teachers, the researcher could understand the difficulties of teachers in planning and developing a test for learners with varied abilities. The investigator identifies a number of studies related to individualized instruction but, the studies related to individualized evaluation strategy are less. Utilization of technology in education especially in evaluation is relatively meager as compared to information and communication technology.

Hence, the researcher is intrinsically motivated to begin a study on Computerized Adaptive Test (CAT) development, which is a type of CBT, to assist teachers at different levels and to develop testing environments that encourage active interaction. Hence, the present study is a felt need for the present social context and a collective study of few dimensions that essentially contribute to the effectiveness of Computerized Adaptive Test package in physics is felt worth an attempt.

1.5. PROBLEM EXPLANATION

The present study entitled “Influence of Structured Computer Adaptive Test and Potentials on Response Pattern of Testees in Individualized Physics Evaluation” aims at developing and validating a computer Adaptive Test material for school students to estimate their knowledge and the level of clear understanding in Physics concepts and establishing the adaptive ability of the test materials on basic concepts in physic to the learner’s internal potentials and external environment facilitated by computer.

According to the Item Response Theory (IRT), answering ability, when challenged may enhance interactivity, if required internal potentials and externally
favorable environment facilitates. Individualized response will provide an uncontaminated, isolated environment for perfect monitoring, when using CAT as a platform for evaluation process in Physics. The depth of knowledge of the learner by challenging their abilities and estimating their potential needs calibrated item bank as proposed by Weissman (2003); Wauters, et.al., (2010). The tool has the potential to grade the testees to various levels and instigate the learners to do the best of their performance by drawing out the best in the individual. In this process of answering, the potentials that operate can play the role of determinants of achievement rate and level of the learners. The potentials could be internal or external in nature and may be operating individually or collectively. Individuality is expected in the nature of performance, but certain abilities are assumed in general and the degrees of influence will be different.

The influence of the package is observed in terms of the amount of interactivity caused on the learner and the duration of interaction or Response Time, which may be decided by the knowledge and skills possessed by the learners. The nature of interactivity and the Response Time may be the indicator of the effectiveness of the Test mode package as revealed by the test score. (Computer Adaptive Test score-CAT score). Thus the Computer Adaptive Test package is used as a means to measure the adaptive ability of the CAT.

1.6. SIGNIFICANCE OF THE STUDY

National Assessment and Accreditation Council, India, calls for technology intensive approach to improve the evaluation system. Computer technology can open up opportunities for developing innovative assessment tools in almost all the fields of education, especially in science education (Lina Ashar, 2012). The nature of computers as information processing tool, user–friendly, interactive learning
environment and the possibility of designing instructional tools to meet individual needs of students, make computers potentially powerful tool for assessment. Computer Based Testing (CBT) could be defined as any type of assessment that is administered through the computer.

Computer based testing provide instant feedback to the teachers as well as the individuals. So self-evaluation is generally automated. It avoids bias or subjectivity and poor communication, (Gillis & Griffin 2005). In this the examinee is free from subjective influence and was given opportunity to utilize, his true potential and he is free from fear and tension. When the examinee is administered a test via the computer, the computer can update the estimate of the examinee’s ability after each item and then that ability estimate can be used in the selection of subsequent items. It assists teachers in categorizing the students in to different level (high achievers slow learners, of average scorers). It encourages active interaction and develops high level thinking skills, such as and problem solving, which are very much useful in science education. Due to the various significances over the traditional test, it is intended to develop new technology oriented tool for evaluation.

Computer Adaptive Test (CAT) is a different kind and a type of CBT, a handy tool and an assistive device with inbuilt provisions for monitoring and routing various associated functions automatically and providing evaluation according to the ability. Effective self assessment is only useful if, after the self assessment there is an opportunity for revision for further improvement. CAT can be used for a wide variety of purposes.

CAT-report conveys to students about their performance, they can use the result of their self assessment for further improvement (Moss & Brookhart 2009). It also help the teacher to guide the learners by identifying the ability, experience and background
of their students, monitoring their learning and providing feedback to them. It helps in identifying whether an individual has met the specific objectives of a course.

The basic idea of CAT is that test items are selected by the computer to individually match the ability level of each student. In this manner, the test is tailored to each student. The nature of responses is graded and evaluated with reference to various norms. It indicates individual’s level of achievement in a particular domain.

CAT has the maximum degree of adaptively, since; they can be adapted for each examinee based on the amount of difficulty and order in which the items are administered to each examinee and so there is not a unique test pattern for all examinees. In CAT, however, each examinee can potentially see a different collection of operational test items that are selected by them, from a large pool of calibrated test items.

CAT helps in the efficient estimation of an examinee’s ability and to classify the examinee’s in to different categories without allowing them to cheat on the test and it is very much effective in reducing the bias of the ability estimates. It also aims at diagnosis and to provide information about specific content areas in which an examinee needs. This simple tool of assessment is raised qualitatively to assume importance, when it is empowered to grade the testees at various levels and instigating to do the best of the testees’ performance. It draws out the best available potentials that are buried in an individual.

Many researchers have proved that the effect of CAT is very positive. It helps to develop inquiry skills, scientific knowledge and scientific reasoning skills. It clears misconceptions and enhances the learners self esteem. It is very useful for the lower achieving students and develops interest in science. It is essential for evaluating large groups of relatively homogeneous individuals and for administering diagnostic testing.
Students learn the material faster and this helps for longer retention. It also has the potential to create a conductive atmosphere to active evaluation. It develops motivation and self esteesms by allowing students to take more responsibility for their learning, because of the immediate feedback they receive.

Teachers have a great role to play in meeting the needs of pupils from a large number of diverse backgrounds and increasingly diverse special needs. ICT technologies like computer Adaptive testing, allows students to enjoy examination with less stress and anxiety which are less impacted by time restrictions, learning schedules etc. CAT can thus be effectively implemented to all kind of learners for enhancing their academic performance, indicating an individual's level of achievement in a particular domain. It also aids with the transmission and administration of marks which can be automatically entered into information management systems and student records databases.

So it is imperative to adopt such a testing tool for quality evaluation and time reduction. But, in spite of the repeated requests of all these commissions for reforming examination, the education system has made comparatively very little progress for the last two decades. The explosion of knowledge together with the electronic revolution has contributed very little in an average Indian school. It is rather primitive in its approach; much has yet to be done. Hence, by suitably adapting the advantages of modern tools such as ‘Microsoft. Net’, the researcher developed a CAT package, which can judge student’s performance from the low to high levels of cognitive complexity and helps in ability estimation. Individuality is expected in the nature of performance but, certain abilities are assumed in general, and the degrees of its influence are verified. The potentials could be internal or external in nature and may be operating individually or collectively. The flow of responses set the trend, which must be able to
project the future and predict the caliber of the examinees. The response study made by the researcher helps in classifying the examinee into different levels.

1.7. COMPUTERIZED TESTING

1.7.1. Technology in Monitoring

The designing of “Innovative Items Structuring” is an emerging area within the field of computerized testing. These items are generally characterized as utilizing features available through the computerized medium that are hardly available in paper-and-pencil format, which result in the creation of a new testing environment. The application of computers to evaluation is not new but as old as its inception, but the need for mapping evaluation behaviour was long felt from the time of Anastasi (1968) in the field of educational testing. According to cognitive science, computers are information-processing tools with a capability to be developed into thinking tools (Rowe, 1993; De Mey, 1992). It can also develop the problem solving abilities in an individual (Pol, Harskamp, 2009). Tracking the progress of the testee and keeping, a cumulative record of progress individually is generally difficult, but technology has become a handy tool and assistive device with built in provisions for monitoring and routing various associated functions. Giving a test, scoring and attributing weight is a common task in use from formative evaluation to online assessment. This helps to replace the monotony of paper pencil assessment and subjective testing.

Computer-assisted test or Computerized Adaptive Test is prevalently in use from classroom application to research tool. The role expectation of the software is to receive the response and verify with the keys for the answers. The end result is accepting or rejecting. The questioning can be at any range from simple to higher order according to the previous performance. To add more dimensions, the software can
assign weightage to the response in terms of attempts. Further, the nature of response is graded and evaluated with reference to various norms.

1.7.2. Models of Adaptive Test

Anastasi (1982) has proposed various modes of computer application in testing for classification. Computer-based testing (CBT) could be defined as any type of assessment that is administered through the computer. However, computer-based testing can encompass many forms, depending on how adaptable the test is on the learner level. For example, some CBT, which are also called computerized fixed tests, are purely linear (Parshall et.al., 2002). These tests most closely resemble paper and pencil tests, since they are in fixed form, fixed length, and the test items are organized in advance and placed in a predetermined order. In contrast to computerized fixed tests, computer adaptive tests (CAT) are the computer-based tests that have the maximum degree of adaptivity. Since they can be adapted for each examinee, based on the amount, difficulty and order in which the items are administered to each examinee. Assessment applications for computers can be broadly classified into two categories: traditional and contemporary. In traditional applications, the infrastructure is rigidly algorithmic. Examples include forced-choice and multiple-choice testing, grading and record keeping. In more applications that are contemporary, the infrastructure is quasi-algorithmic or non-linear in nature. Examples include constructed response testing, adaptive testing, figural response testing, simulations, and solution pathway analysis (Kumar David, 1996).

Today, powerful microcomputers are not only beginning to affect a redesign of the structure and content of school curricula and the entire process of instruction and learning, they also have a decided impact on the types of tests created and used to assess that learning. In fact, computerized testing is increasingly being viewed as a
practical alternative to paper-and-pencil testing (Kingsbury & Houser, 1993; Clariana Roy et.al., 2002; Georgiadou, et.al., 2007). Tests administered at computer terminals or on personal computers are known as computerized tests. Given the advantages of individual, time-independent testing, computer-based testing will no doubt prove to be a positive development in assessment practice (Brown, 1997; Weissmen, 2003; Mohanty, 2010; Fehr, Charles, et.al., 2012).

1.7.3. Computer Adaptive Test (CAT)

In the 1960s and 1970s, the U.S. Department of Defence perceived the potential benefits of adaptive testing and supported extensive theoretical research in CAT and Item Response Theory (IRT). IRT is based on probabilistic theory; that is, it calculates the probability of a given person getting a particular item right (Alderson et.al., 1995; Samejima, 2000). Examinees' scores and item total statistics are transformed into one scale so that they can be related to each other. If a person's ability is the same as the difficulty level of the item, that person has a 50-50 chance of getting that item right. If their ability is at a lower level than that of the item, that probability decreases. As the level of students' ability increases, so does the probability of a correct response (Alderson et.al., 1995; Reese Lynda. 2003; Wu, Huey-Min., et.al., 2012).

Early attempts to build adaptive tests by the U.S. Army, Navy, and Air Force were often less than successful, very expensive, and used large-scale computers. However, by the early 1980s, personal computers had acquired the power of the large-scale computers of earlier years, and the pioneering efforts of IRT theorists had perfected the psychometric model underlying CAT. In the late 1980s, CAT finally moved out of the realm of theory and supposition into the sphere of possibility and implementation with the advent of the College Board's CAT Graduate Record Examination and with the work of in-house researchers in foreign language education.
at the Defense Language Institute and at universities throughout the United States, Britain, the Netherlands, and other countries.

Today, with software development companies assisting test developers with their own institutional CATs, computer-based testing has finally become a viable alternative to conventional paper-and-pencil testing. It is anticipated that in the future, more and more computer companies and academic institutions will jointly producing testing shells that can be used to create CATs for placement, achievement, and licensing purposes. Weiss & Kingsbury (1984), Spray & Reckase (1994), Reese (1999 et.al and more recently He, Wei, & Reckase, (2008) suggests CATs for situations where the main interest is in estimating the ability of an examinee, than to classify the examinee into two categories, e.g., pass-fail, master/non-master.

In Computerized Adaptive Testing, the computer tailors a test sequence according to an examinee's level of achievement and ability. For example, based upon the kind of response made to a question, the computer will review and decide whether to stay on the same level to ask another question or clarify background knowledge, or proceed to a higher level question and to a different topic (Welch & Frick, 1993; Chatzopoulou, & Economides, 2010). In Item Response, either the test may be used as routing test (Sheikan, 2001). It is more applicable for norm - referenced purposes like general proficiency testing (admission). In CAT, since everyone takes every item, all examinees are administered test on each concept.

It is important to note that not all these applications are the sheer result of the availability of computer hardware but of the introduction of efficient programming methodology.
1.7.4. Estimation of Ability

From paper-pencil format of evaluation, there is a gradual transition to computer on-line testing being employed in arena of international employment (Lawrence, 1998). However, such developments have not made any impact on the present evaluation system and traditional test formats prevail in all. Still the examinee has not been free from the subjective influence and never given opportunity to utilize his true potential that’s what is needed in both criteria and in norm references with the maximal expectations. Paper-and-pencil tests are typically "fixed-item" tests in which the examinees answer the same questions within a given test booklet. Since everyone takes every item, all examinees are administered with items that are either very easy or very difficult for them. These easy and hard items are like adding constants to someone's score. They provide relatively little information about the examinee's ability level.

When an examinee is administered a test, via the computer, the computer can update the estimate of the examinee's ability after each item and then that ability estimate can be used in the selection of subsequent items. With the right item bank and a high examinee ability variance, CAT can be much more efficient than a traditional paper-and-pencil test. With computer adaptive tests, the examinee's ability level relative to a norm group can be accurately estimated during the testing process and items can be selected based on the current ability estimate. Examinees can be given the items that maximize the information (within constraints) about their ability levels from the item responses. Thus, examinees will receive items that are very easy or very hard for them. This tailored item selection can result in reduced standard errors and greater precision with only a handful of properly selected items. Thus, a Computer Adaptive Test (CAT) is a test customized by a computer system based upon the responses level of the testee (Wainer, 1990; Wang et.al., 2003)
The basic notion of an adaptive test is to mimic automatically what a wise examiner would normally do. Specifically, if an examiner asked a question that turned out to be too difficult for the examinee, the next question asked would be considerably easier. This approach stems from the realization that we learn little about an individual's ability, if we persist in asking questions that are far too difficult or far too easy for that person. We learn more about an examinee's ability when we accurately direct our questions at the current level of the examinee's ability (Wainer, 1990; Weissmen, 2003; He, W., & Reckase, 2008). Thus, in a CAT, the first item is usually of difficult level for the test population. An examinee who responds correctly will then move to the next difficult level question. An examinee, who misses the first item, will be given an easier level question. Therefore, it goes, with the computer algorithm adjusting the selection of the items interactively to the successful or failed responses of the test taker. So computer adaptive tests (CAT) are the computer-based tests which are created and adapted specifically for each examinee, based on the examinee’s ability estimate, and on the way in which each examinee has responded to the previous items that have been administered Huo, Yan, 2009; Samejima, Fumiko, 200; Yen, Yung-Chin., et.al., 2012).

Weissmen (2003), Yen, Yung-Chin., et.al. (2012) suggest that the adaptive test can follow a wide variety of procedural model using the computer and the level of the testee is automatically selected in consultation with their earlier performances. These easy and difficult items are like adding constants to some one scores. They provide relatively the information about their ability levels. In CAT, the students are channeled by a routing test into items exactly at their ability levels. In essence, the computer uses the information it gets from the routing test to select items specifically for each student's level of ability. Each student essentially takes a different test. CAT requires
a) a large item bank be piloted and analyzed,

b) the developer have background in item-response theory statistics, and

c) the test developer have considerable knowledge of computer programming (e.g., Visual basic, Micro soft Dot net etc.) or internet browser programming.

1.7.5. Item Response Theory (IRT)

Item response theory (IRT) also known as latent trait theory, strong true score theory, or modern mental test theory, is a paradigm for the design, analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other variables. The name ‘Item Response Theory’ is due to the focus of the theory on the item. IRT is based on the idea that the probability of a correct/keyed response to an item parameter, on which items are characterized include their difficulty (known as "location" for their location on the difficulty range), discrimination (slope or correlation) representing how steeply the rate of success of individuals varies with their ability (Mullins, et.al., 2011). The purpose of IRT is to provide a framework for evaluating how well assessments work, and how well individual items on assessments work. The most common application of IRT is in education, where Psychometricians’ use it for developing and refining exams, maintaining banks of items for exams, and equating for the difficulties of successive versions of exams. Item response theory (IRT) seeks to model the way in which latent psychological constructs manifest themselves in terms of observable item responses; this information is useful when developing, evaluating, and scoring tests. IRT gives a detailed description of the performance of individual items, and measures the response-profile quality in computer-adaptive testing, which can dramatically reduce testing time (Froehlich, Amy Foodwin, 2000). With the rapid developments in computer technology and measurement theory, and by using an adaptive testing strategy, an assessment can be
accomplished in less time, provide more accurate and reliable scores. Estimating an examinee’s ability has become a relatively easier task with the advances in Item Response Theory in conjunction with advances in computers (Betz, Nancy & Turner, Brandon, 2011).

Often Computer Adaptive Testing has been applied to a testing process that makes item selection decisions after each item is answered. That is, a CAT uses each item as a Routing test. Measurement precision is consistently refined after each item is presented and answered. This method is more efficient than the traditional paper-pencil test and therefore CAT will provide results with higher accuracy (Georgiadou, et.al., 2007).

1.7.6. Application of Item Response Theory

Tool of test from simple Criterion Referenced Test and Norm Referenced Test has now be elevated to invariance by application of Item Response Theory (Reed, 1997). Item discrimination is an index of item’s ability to discriminate among examinees, but invariance is the concept that indicates item parameters and ability estimates that are stable regardless of population. Hambleton & Swaminathan (1985), Weissmen (2003) and recently Wauters, et.al. (2010) proposed IRT model that evolved into estimating examinees ability by the item, which are pre-calibrated and edited to fit into the Computer Adaptive Test (CAT). Reese et.al (2003) used two-stage testing, which was later developed by Pol, Henk et.al (2009), based on the Two Stage Testing design of James Dean Brown (2001). It is especially important that superior items be used on the routing test, which means the items must be well written, must vary considerably in difficulty, and must discriminate very well. If the items in the routing test are not working particularly well, then their ineffectiveness in channeling students
into the measurement tests could create considerable unreliability in the resulting scores.

Under Item Response Theory, plenty of research works have been done and using them as base, the Researcher has built up the present study on Computer Adaptive Test with three measurement levels. The researcher has designed a new CAT in Physics concepts. The purpose of the test was to estimate the ability of the testees’ through their movement over three different levels of questions i.e. difficult, average and easy levels. The researcher used the parameters Difficulty Index and Discrimination Index at the preliminary stage of item validation in order to refine the items of the Item Bank.

1.7.7. Performance Mapping

Graphical Mapping is simple but proposed to be a dynamic method that facilitates easy access in documenting or replicating richly constructed student responses. Designing a sequenced set of student activities and constructing probable maps associated with each learning activity reveals divergence in thinking on the encounter with response demand. Performance mapping is an approach to documenting the behavioral pattern in computer based interaction. Thus, the purpose of the software

![Diagram of CAT Performance Chart]

Fig. 1.1. Diagrammatic Organization of CAT Performance Chart (Adapted: Anastasi, 1984)

mapping and its interpretations, the researcher could arrive at certain new types of
response patterns depicted by various response curves. In this study, an attempt has been made to study the learning responses of the learners (during the administration of the CAT package) with graphical representation in order to see the evolution of their pattern of response curves.

1.8. CAT PACKAGE DEVELOPMENT

Using pyramidal testing model the researcher has designed the tool of research on the basic concepts in physics subject. A multiple chain of items is developed as a norm referenced with wide range in a taxonomical level. In this case, all the examinees begin with an item of highest difficulty. If an individual’s response to this item is correct, he or she is routed upward to the next question of more difficult item; if the response is wrong, he or she moves downward to the next average level item. In the easier level item, the number of attempts made by the testee is noted and marks are reduced for each attempts This procedure is repeated after each item response, until the individual has given response to all the items in the test. In this method, learning of the concepts will also takes place. For testing to be fully adaptive, the researcher has given due considerations to the following points;

a) A delivery system that is able to retrieve items quickly from a large pool (450 items) and display them to the examinee.

b) A rule for selecting the next item to be displayed as a function of the response on the previous item.

c) A rule for scoring the ability of the examinee, based on the responses to the items displayed, after the completion of the test.

In this CAT, the examinee takes the first item, which identifies the examinee’s ability to face the test like the norm referenced test and routes him to three levels of interactivity as difficult, average and easy. Each of the total number of 450 items (150 X 3 levels) has a potential to test a learner according to their ability where in; the items
are meant to establish this ability like criterion referenced test. The capacity to elicit response must be in built in the Test Tool and the researcher has taken strenuous efforts, consultation and field trials to build up the Question Bank to meet these demands. Technology has been handled appropriately to provide the required support in assessment, classification and routing the testees according to their performance rating or ability level.

In the development of the computerized test package, the scientific principles in the phases of its development and the procedures of validation are suitably identified and applied. The procedures of development and validation are unique and the indicators of their effectiveness are not the achievement in CAT (Computerized Adaptive Test) alone, but it is found to be the Learning Time and number of attempts as well. The extent to which the external variables like locality of the Institution (rural/urban), the Level of Education, Gender, different types of Learning environment in school (state board, metric board, central board), were found to interact with internal variables like Aptitude in Physics, Attitude Towards Technology, and Attitude Towards Individualization to be the determinants are attempted.

The computerized test package developed is the base for computer adaptive test and some of the basic concepts in Physics are identified to construct multiple-choice questions that are specially designed to have three levels like – the high, average and low. In the adaptive test, each examinee is exposed to a difficult level question and depending upon their performance, they are directed to the next question of high difficulty or to the same question at low difficulty. Thus, each person completes the entire test comprising of 150 items, according to their own pace, but predetermined by their potentials to follow different paths. The ground for test is constructed on some of the basic concepts in physics, which forms the core for higher studies. In the computer-
adaptive test, each question is shown on a personal computer screen one at a time. On the test, questions are set at high, medium, and low difficulty levels. The first question on a test is of high difficulty; the relative difficulty of the next question depends on the learner’s answer to the first question. If the learner answered correctly, the next question will be of greater difficulty. If the answer was incorrect, learner will be directed to the same question with less difficult level, and so on. However, the choice of subsequent level of the questions is not only based on whether the preceding answer was correct or incorrect, but also on the difficulty level of the preceding question. This procedure is repeated for each of his answers. In this way, the CAT adjusts questions to the ability of the learner.

LEVEL 1: The first level is challenging the learner’s critical awareness and analysis by provoking him to think at abstract level. The question item requiring application, analysis and synthesis on the part of the learners are included at this level.

LEVEL 2: The item at average level considers the general level learning ability of an average student. It is neither too difficult nor too simple, but mostly at knowledge and comprehension level. However, the higher ability requirement is expected from the learner so that, he can perform critically at this level.

LEVEL 3: The items designed under this category in general are simple in terms of content recollection. The answers provided are direct and guiding in nature. The key alternatives are contrasting and the distraction level is low.

1.8.1. The Computer as a Means

The researcher has developed a computerized adaptive test package that contains 450 questions arranged in a specific pattern. But no student will get the exact pattern in which the questions are organized and presented, and each individual will follow a different path depending upon his level of performance. Thus in this study the
computer is used not as a tool to supply questions, but as a means of evaluating the knowledge level by sequencing the presentation individually according to the learners’ abilities. Along with testing on various cognitive levels, this test provides freedom to choose his challenging level. An examinee’s upward or downward movement in the sequence is based upon his response in each of the previous item. The format of the test motivates the learner to move to the next higher level, as if two questions are answered correctly then next higher order question with proposed level of difficulty is exposed to the testees.

1.8.2. Potentials of the Present Computer Adaptive Test Package

The CAT package prepared by the researcher has the following advantages, When an examinee is administered the test via the computer, the computer can estimate the examinee's ability after each item and then that ability estimate can be used in the selection of subsequent items. As the item bank contains large number of questions, CAT can be much more efficient than a traditional paper-and-pencil test. Paper-and-pencil tests are typically "fixed-item" tests in which the examinees answer the same questions within a given test booklet. They provide relatively little information about the examinee's ability level. But in CAT since everyone takes every item, all examinees are administered a items that are either very easy or very difficult for them. These easy and hard items are like adding constants to someone's score. Consequently, large numbers of items and examinees are needed to obtain a modest degree of precision. With computer adaptive tests, items can be selected based on the current ability estimate. Examinees can be given the items that maximize the information about their ability levels from the item responses. Thus, examinees’ will receive few right items that are not very easy or very hard for them. This tailored item selection can result in greater precision with only a handful of properly selected items.
Tests are individually paced so that an examinee does not have to wait for others to finish before going on to the next section. Self-paced administration also offers extra time for examinees who need it, potentially reducing one source of test anxiety. Significantly, less time is needed to administer this CAT than fixed-item tests since fewer items are needed to achieve acceptable accuracy. CAT can reduce testing time by more than 50% while maintaining the same level of reliability. Shorter testing times also reduce fatigue, a factor that can significantly affect an examinee's test results. This CAT cannot accommodate an option for examinees to review their answers, Wainer's (1990) testlet approach is a notable exception. If review and answer changing were possible, a clever examinee could intentionally miss initial questions. The CAT program would then assume low ability and select a series of easy questions. The clever examinee would then go back and change the answers, getting them all right. The result could be a high percentage of correct answers, which would result in an artificially high estimate of the examinee's ability.

Because examinees may be of high or low proficiency levels, the CAT is designed in such a way as to provide adequate assessment for the entire range of ability represented in the examinee population as suggested by Green et.al., (1995), Wise, Steven et.al., (1997) and Ogan, Amy, et.al., (2009). A pre-test – post-test, quasi-experimental design is adopted for this purpose.

1.9. VARIABLES OF THE STUDY

In the present study, Computer Adaptive Test is the assessment platform which encompasses several dependent variables like Computer Adaptive Test score, Response time, Attempt, and Mean Achievement Score. The following variables are measured when the ability of the learners is estimated through CAT package on Physics.
1.9.1. Dependent Variables

1) CAT Score is the score obtained by the learner through the interaction with the CAT package in the form of the ability estimate. In this the examinee takes the first item which identifies his ability to face the test and routes him to three levels of interactivity as High, Average and Easy. The total number of 450 items (150 x 3 levels) has a potential to test a learner according to his ability, where in the items (three in numbers) is meant to establish this ability like criterion referenced test. The maximum score achievement is 1500 (All 150 Higher level questions x 10) and the lowest achievement possible is 150 (All 150 easy level questions x 1).

2) Response Time is the interaction time of the learner with the testing package. It is the time taken by the learner to answer all the 450 questions of the Computer Adaptive Test package, and it is measured internally by the software itself in minutes automatically and can be seen or displayed in the report sheet once all the questions are answered. Response time is the actual test time the examinee spends with the item which is monitored by the researcher continuously as interactive time.

3) Learning Rate is in parity with the spirit of involvement in the process. The pace of interaction is dependent on learning kind, material, environment and the process associated. It establishes the behavior of the examinee in terms of the testing condition provided by the structured environment in achieving the objective. It is calculated by the Response Time of interaction of learner with the CAT package. It is the time taken by the learner to complete the various test items of the CAT package, which is calculated in minutes. Since the researcher has employed the principle of ‘self pacing’, the testee is allowed to take as much time as he needs as his interaction time.
4) **Mean Achievement Score** measures the gain in the knowledge of the learner as a consequence of his interaction with the test material. It is the mean achievement score of the learners in the criterion test which is given by the difference between the mean Post-test score and the mean Pre-test score and is being labeled as the product of the learner’s interactions with the questions and learning potential.

5) **Attempt means** response of the learner, which often reveals the nature of learning by indicating the successful learning or confusion or unlearning. The unsuccessful learning is revealed by the wrong response but at times it can couple with certain prodigal behaviors’ of cheating in computer and marked as ‘Attempts’. The nature of instructional material is measured with reference to learning time and its number of or frequency of usage(attempt). Thus, attempt is used as an essential parameter in the process of validation of the instructional frame, the learning ability of the learner and the instructional ability of the software.

1.9.2. **Independent Variables:**

Independent Variables are of two types.

A. External Independent Variables, and
B. Internal Independent Variables

A. **External independent variables:**

A rationale for choosing the above variables is discussed here with.

1) **Level of Education** denotes the subject contact (basic concepts understanding and retention in Physics) of the learners of different levels. In this study school students and physics undergraduate college students have been selected from three different types of school back ground, as a result of which there is difference in their previous knowledge of the subject matter. The school students
are continuously exposed to the basic concepts in their course of study while the college students have the residual knowledge of the basic concepts in their course of study.

2) *Gender* refers to the sex of the learner comprising of boys and girls. As the psychological conditions, exposure to society, experience with technology and computers vary differently, there may be a significant difference in their mental development, learning potential and attitude formation. Hence, there may be a difference in their aptitude to Physics, attitude towards technology, attitude towards individualization and technology preference. Since the above mentioned factors will have their influence upon the computer technology and the adaptive testing material, a comparison between boys and girls will reveal the impact of gender in ability estimation.

3) *Educational Environment* discriminates the learners from three different types of school. The schools provide different activities and follow different syllabus and course materials to deliver the same subject. These factors will have their influence on understanding and retention of the basic concepts, thus the students of different school environment will get multiple opportunities to learn the subject over and again. The academic performance rate of central school students may be better at higher level as compared to the state and matriculation students. The difference is because of the different methodology and the hand on experience provided at higher secondary level. Hence, educational environment has been considered as one of the external variables.

4) *Locality of the Institution* indicates the two important demographic parameters such as rural and urban situation of the school atmosphere. The researcher is
curious to find out the influence of these two parameters because there may be
difference in the learning potentials and internal attitudes of the learners of these
two situations.

5) *Level of Computer Knowledge* indicates the familiarity with the computer, this
parameter is included as this may influence in the learners while working with
CAT. This evaluation tool uses computer as a platform, so the difference in their
exposure or familiarity will definitely have an impact upon their interaction with
the computer which intern may affect the CAT score.

**B. Internal Independent Variables:**

1) *Technology Preference* considers the different preferences of learners among the
students of different school environment and different locality of institution, in
terms of their knowledge and understanding of technology. One category of the
students may expose to the various tools of teaching and their applications where
as the other category students may not have such experiences with the teaching
tools. This difference in their exposure will definitely have an impact upon their
technology preference which in turn may affect the CAT score.

2) *Attitude towards CAT*, during the interaction with CAT denotes the learner’s
positive or negative opinion regarding the use of technology in education and
individualized instruction. It was assessed by the teachers’ after observing the
testees’ performance during the execution. Male learners have pronounced trend
of using computer for school – related applications prior to Female learners.
These attitudinal differences will definitely affect the ability estimate of the
learners.
3) **Aptitude in Physics**: A natural or acquired disposition or capacity for learning Physics or readiness in learning Physics is known as Aptitude in Physics. This aptitude is an essential factor in all kinds of instructional and testing processes (Slykhuis David & Park, John, 2006). Physics deals with the study of the basic Laws that governs the natural world and their manifestation in different phenomenon. Physics Aptitude Test (PAT) is a non-verbal, multiple choice test in Physics concepts Developed, validated and adapted to various situations by the researcher. The questions in this need the practical skills. The pictorial representations of different aspects of the basic concept are used to assess the cognizant of the students in the subject. These questions connect the theoretical concepts in physics to experience in the real world of application. It is used to measure the Physics Aptitude in the examinees to identify their ability component for facing the test items.

4) **Attitude towards Individualization** refers to the learner’s positive or negative opinion about individualized instruction. Female have pronounced trend between computer usage for school related applications prior to college and preference for technology based lecture tools during college (Kate Barber & Melissa Maquilan, 2004). These attitudinal differences will definitely affect the ability estimate of the learners. Given its increased use, it is important to understand how instructional technology and the technologically rich environments are influencing student attitudes toward learning. An important aspect in successfully implementing instructional technology is user acceptance, which may be influenced a great deal by users’ attitude.
1.10. OBJECTIVES OF THE STUDY

The following objectives were framed for the study:

- To integrate technology in self-assessment process at school level in Physics.
- To organize question bank with item difficulty level suitable to different categories of learners.
To develop a CAT package in the selected basic concepts in physics at school level.

To find out the effectiveness of the prepared Computerized Adaptive Test package in terms of
- Number of attempts made by the learner.
- Average response time taken per student.
- Achievement score of the student in CAT.

To study the comparative effectiveness in achievement through CAT on different types of learners high, low and average achievers in school.

To assess the adaptive ability of the CAT package in computer based evaluation.

To correlate internal and external potentials with response pattern of learners.

To measure the learning rate of learners.

1.11. ORGANIZATION OF THE THESIS

The thesis consists of five chapters:-

Chapter I: Introduction, explains briefly, the work done by the researcher regarding the present study. It also explains the need, significance and how technology can be effectively utilized for evaluation purposes.

Chapter II: Review of literature, gives an elaborate view of the literature study made by the researcher regarding the present work.

Chapter III: Method, Procedure and Collection of Data, discusses the methodology adapted by the researcher for the present study. It also contains the details of the CAT package development, and also the procedure for the development and validation of the tools used in the present study.
Chapter IV: Analysis of Data and Interpretation, gives an elaborate detail of the data analysis and Interpretation done in this research to arrive at a conclusion regarding the present study and also to find out the effectiveness of the prepared CAT package.

Chapter V: Conclusion, Findings and Suggestions includes the details of the findings arrived, while testing the hypothesis. It also explains the implications of the present work.

1.12. DELIMITATIONS OF THE STUDY

The following are the delimitation of the study;

(i) The study was limited to the sixty five students of few selected schools in Madurai District, because each individual have to be provided with provisions for CAT package and the system with other facilities, which has taken a considerable period of time and space.

(ii) The demographic variables such as parent’s educational qualification, parents salary, previous learning environment, which are essential and can influence the learning are not considered.

(iii) The higher level software technology was not adapted since, the hardware facility available in the institutions were not permitted.

(iv) Animations and pictures are not used for the presentation of frames it could have been used in the text in each frame in order to include the component of variety.

(iv) In built provision for calculating the Learning rate was not programmed in the package.

(v) Among the various potentials, the study is limited to the potentials like Attitude, Aptitude, Technology preference and Individualization.
1.13. SCOPE OF THE STUDY

The study was undertaken with the hope that it will be beneficial to the system of education, the teachers, the learners and the educational technologists as it aims for the development of instructional software, which can suit to different kinds of assessment techniques, adopted in question bank, cultural assessment and performance design. Further, the educational technologists from other disciplines may adapt the application of the unique techniques and procedures employed in the preparation and validation of instruments to similar situations. This kind of study will encourage more teachers to follow innovative methods of evaluation in individualized instruction, CAI etc. The learners will also be motivated in self-learning and self-assessment of knowledge on any discipline as it provides for the individual differences. The results of this study will throw more light on testing mode in Physics through CAT which helps us to realize the importance of the utilization of computers The measurement of learning time is attempted along with achievement as an index of effectiveness of CAI package as suggested by earlier studies, to find out the nature and process of learning in individualization of instruction. The selected variables studied are expected to provide the base to identify the nature of interaction between the related variables on CAT.

1.14. CONCLUSION

For quality enhancement in the present education system, the first step is to find changes in the traditional evaluation system. Expanding the access to higher education and ensuring the quality are the aims of our government. By developing suitable computer based testing tools, the evaluation will be able to identify potentially strong and knowledgeable persons. Continued research and development efforts are needed in order to use computers as tools that are unquestionably effective to trace any behavioural pattern in assessment of science education.
In the present study, opportunities are provided for the learners’ for performing at various levels and the opinion is given to the individuals with freedom to choose according to their self-estimation of ability and risk taking nature. Each test takers’ will follow a different path, depending upon their performance in the test. Each learner’s path is unique as no body can guess or modify one’s path of progress while using the computerized adaptive test package. At an operational level, the criterion fixed in the whole process is the ability of the test item to estimate the residual potential of the testee to the maximum. It provides not only a challenging situation, but also instigates the testee to perform at the optimal level (Lina Ashar, 2012). When a generally low profile testee is able to achieve high (Brecht, H. David, 2012), the conclusion is favorably skewed to the ability of the items than the expressive ability of testee. The whole operation is thus reversed in the test; where the ability to respond to a test item is estimated than the testee is measured in terms of an established norm. There is a graded variance created, by the test item and the respondent collaboratively. Hence, the hybrid nature of the test item is because of the principles used in construction of the test items and the operational outcome expected in the experiment. The items have a different ability, which is able to reach the residual potential, which is not having a chance of expression until then. It may be believed that the potentials with other variables necessary for responding to the item have been utilized to the utmost available in the testees. Given a freedom to express and opportunities to interact, the ability can be expressed automatically, concludes the researcher. The next chapter deals with Review of related literature.