CHAPTER - II

Review of Related Literature
## CHAPTER II - REVIEW OF RELATED LITERATURE

### SECTION I

<table>
<thead>
<tr>
<th>2.1.</th>
<th>Introduction</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.</td>
<td>Researches Related to Computer Assisted Instruction</td>
<td>35</td>
</tr>
<tr>
<td>2.2.1.</td>
<td>CAI Package in Science</td>
<td>36</td>
</tr>
<tr>
<td>2.2.2.</td>
<td>CAI for Achievement &amp; Longer Retention</td>
<td>38</td>
</tr>
<tr>
<td>2.2.3.</td>
<td>CAI Effectiveness</td>
<td>42</td>
</tr>
<tr>
<td>2.2.4.</td>
<td>CAI Meta Analyses</td>
<td>46</td>
</tr>
<tr>
<td>2.3.</td>
<td>Researches Related to Computer Adaptive Testing</td>
<td>50</td>
</tr>
<tr>
<td>2.3.1.</td>
<td>Computerized adaptive test development</td>
<td>50</td>
</tr>
<tr>
<td>2.3.2.</td>
<td>Procedures of CAT, IRT</td>
<td>54</td>
</tr>
<tr>
<td>2.3.3.</td>
<td>Test designs in CAT</td>
<td>63</td>
</tr>
<tr>
<td>2.3.4.</td>
<td>Effectiveness of CAT</td>
<td>70</td>
</tr>
<tr>
<td>2.3.5.</td>
<td>Gender difference in CAT/ CAI</td>
<td>78</td>
</tr>
<tr>
<td>2.3.6.</td>
<td>Students’ Attitudes Towards CAT</td>
<td>82</td>
</tr>
<tr>
<td>2.3.7.</td>
<td>Learning Time</td>
<td>89</td>
</tr>
<tr>
<td>2.3.8.</td>
<td>Learning Rate</td>
<td>92</td>
</tr>
<tr>
<td>2.4.</td>
<td>Conceptual Frame of Reference</td>
<td>93</td>
</tr>
<tr>
<td>2.5.</td>
<td>Summary of the Review</td>
<td>97</td>
</tr>
<tr>
<td>2.6.</td>
<td>Conclusion</td>
<td>97</td>
</tr>
</tbody>
</table>
CHAPTER II
REVIEW OF RELATED LITERATURE

2.1. INTRODUCTION

This chapter attempts to survey and critically review studies related to the theme of the investigator chosen for his study. Review of related literature is an important prerequisite to actual planning and then the execution of any research work. Best (1963) writes, “a familiarity with the literature in any problem area helps the students to discover what is already known, attempted methods that are promising and disappointing, and problems that remain to be solved”.

The investigator feels that the study of related literature helps in acquiring information about the studies done in the field, protects against unnecessary duplication, guides in carrying out the investigation successfully and makes the researcher familiar with the steps. Citing studies that show substantial agreement and those that seem to present conflicting conclusions help to sharpen and define understanding of existing knowledge in the problem area, provide a background for the research project, and make the researcher aware of the status of the issue. In the following pages, a review is made on the studies conducted in computer assisted instruction and computer adaptive test related to the present study.

2.2. RESEARCHES RELATED TO COMPUTER ASSISTED INSTRUCTION

This section contains the summary of literature related to Computer Assisted Instruction under the following headings.

- CAI Package in Science
- CAI for Achievement & Longer retention
- CAI Effectiveness
- CAI Meta Analyses
2.2.1. CAI Package in Science

In this sub section studies related to CAI package in science are discussed;


An approach to accelerate inventive preliminary design was studied and this method combines the main advantages of CBR (Case Based Reasoning) and TRIZ (Russian acronym for Theory of Solving Inventive Problem) to transfer physics to industrial technology. Findings showed that inquiry modules were able to facilitate investigation and planning activities in product design stages. It also indicated that these were indeed helping Students to develop their engineering thinking and design skills, increasing their motivation to study (Jou, Min, et.al, 2010).

De Ambrosis, Anna & Levrini, Olivia (2010) carried out a study with a group of high school physics teachers engaged in the Module on relativity of a Master course on the teaching of modern physics. Results showed that the teachers' difficulties in coping with innovative proposals are usually related to the implementation process.

The students in courses that are taught using the Just-in-Time teaching strategy better understand Newton's Third Law after instruction than do students in traditional lecture courses (Formica, Sarah, et.al., 2010).

Selcuk (2011) investigated the pre-service teachers' understanding of and difficulties with some core concepts in the special theory of relativity and found that pre-service teachers at different academic levels have specific and considerable
difficulties with proper time, time dilation, proper length, mass and relativistic density concepts, indicating the implications of new methodologies for physics teaching.

Bandyopadhyay & Kumar (2011) investigated in detail students' understanding of the principle of equivalence. Analysis of responses reveals a number of significant conceptual vulnerabilities in several aspects of the topic and the need for an alternative instructional method. Design and evaluation of an undergraduate final year science communication module for the Science Faculty at the University of East Anglia was done. Students also developed skills relating to self-reflection and how to use this as a tool for future self development (Yeoman, et.al, 2011).

Nilsson, Pernilla & Van Driel (2011) Jan investigated how student teachers perceived the development of their knowledge and attitudes towards physics through video recorded practical workshops based on experiments and subsequent group discussions. The results provided insight into how aspects such as self-confidence and the meaningfulness of knowledge for primary teaching were perceived as important factors for the primary science student teachers' development of subject matter knowledge as well as a positive attitude towards physics.

Anderson, Janice & Barnett, Michael (2011) used video gaming technology to facilitate the understanding of basic electromagnetism with pre-service elementary teachers. Results of this study showed that video games can lead to positive learning outcomes, as demonstrated by the increase in test scores from pre- to post-assessment. Additionally, this study also suggests that a complementary approach, in which video games and hands-on activities are integrated, with each activity informing the other, could be a very powerful technique for supporting student scientific understanding.

Computer-based modeling integrates the interactions of natural phenomenon through the use of models, which provide structure for theories and a base for
experimentation. Utilizing these methods, scientists construct knowledge, and in like manner, students in science construct their understandings in significant ways, addressing their preconceptions and their knowledge of concepts in physics (David Carrejo & William H. Robertson, 2011).

2.2.2. CAI for Achievement and Longer Retention

In this sub section studies related to effect of CAI on Achievement and longer retention are discussed:

“Effects of a Computer Assisted Remediation Program on Basic Skills Mathematics Achievement, Academic Self-Concept, and Locus of Control of Students in a Selected Dropout Retrieval Program in an Urban Setting” was investigated by Reglin, Gary (1988). Results indicated no significant difference on any dependent variable for type of instruction and the interaction between sex and type of instruction. The IMTS program with or without CAI had a more significant impact, by t-test, on males' academic self-concept and locus of control.

“Helping Students To Solve Word Math Problems: CAI as an Alternative Method of Instruction” was studied by Viteli and Jarmo (1989). It was concluded that computer assisted instruction could be used as an alternative to conventional methods of teaching mathematics word problems.

In the study “The Effects of CAI and Hands-On Activities on Elementary Students' Attitudes and Weather Knowledge” Gardner, Catherine and others (1990) determined if a combination of computer assisted, instruction (CAI) with hands-on science activities would significantly enhance students' abilities in the cognitive and affective domains. Hands-on activities appeared to increase scores from pretest to post test.
“Effectiveness and Efficiency of Elaboration Using CAI” by Wilshire, Daniel, (1990) assessed the effects of varying degrees of elaboration in a prose format using computer-assisted instruction (CAI) with accompanying slides on post test scores of college age learners. It was concluded that subjects with high prior knowledge would consistently score better than subjects with low prior knowledge regardless of the kind or amount of elaboration, although both groups would benefit about the same from extra elaboration on tests of increasing cognitive difficulty.

Jegede and Olugbemiro (1991) undertook a study on the title “Computers and the Learning of Biological Concepts: Attitudes and Achievement of Nigerian Students” that compared attitudes towards computer use and achievement in biology for three groups of Nigerian students (n=64): (1) working alone with computer; (2) working in groups of three on the computer; (3) and a control group that received normal instruction (lecture). Students in the second group had the highest scores on attitude. No significant differences were found in achievement.

“The Effects of Computer Assisted Instruction on Student Achievement in High School” by Morrell, Patricia, (1992) presented a study to determine differences in effectiveness of computer-assisted instruction (CAI) compared to traditional instruction. Results indicated CAI tutorials are no more or less effective in promoting student achievement, and the photosynthesis group preferred CAI.

Lwo (1992) conducted a study on “Effects of Individualized Examples and Personalised Contexts in Computer Based Adaptive teaching of Algebra Word Problems” and observed no significant post-test performance differences.

Royo (1994) presented the methodology on the preparation of CAI lesson on the topic rectangles. The study was undertaken to present evidence of a positive effect
on geometry achievement when using a CAI lesson and a significant improvement in the scores was observed.

In a research work titled as “Locus of Control, Self-esteem, Motivation and Problem Solving in 5th grade students using Logo and CAI” Tyler (1994) investigated problem solving. No significant effects or relationships were found to support the hypotheses related to problem solving, locus of control, motivation, or self-esteem. Upon further investigation, all three groups made pre test to post test gains in problem solving and in locus of control.

There was a significant increase in the post test mean from that of the pre test and students have positive attitude toward CAI when Igboko (1994) conducted a study entitles as “NUMBERLAND: A CAI Lesson in Math word Problems”.

“The effects of computer-assisted instruction on the achievement test scores of seventh-grade students” by Strohsahl (1994) revealed that students who received instruction in language, which included CAI, performed significantly higher on the achievement tests in language that students who did not have CAI instruction. Furthermore, male students outscored the female students in language achievement.

Sadatmand (1995) conducted a study on “The effect of Computer-Assisted Instruction on Algebraic Problem solving abilities of Community college students”. Statistical evidence in this study supported the findings that Computer Assisted Instruction (CAI) can enhance student’s intellectual growth in mathematics.

Rader (1996) in the study “The effect of Curriculum Alignment between Mathematics Instruction and CAI Mathematics on student performance and attitudes, Staff Attitudes and alignment” revealed that students who receive computer assisted mathematics instruction would achieve greater.
Caple, Carlous (1996) conducted a study “The Effects of Spaced Practice and Spaced Review on Recall and Retention Using Computer Assisted Instruction”. Findings indicated a statistically significant difference between treatment groups on both recall and retention.

“Effect of Computer Assisted Instruction on Students' Achievement in Global Studies” was done by Adonri, Osa; Gittman, Elizabeth (1998). The assessment found that students using CAI achieved significantly higher scores on a post test than students taught by traditional methods. An attitude survey showed an increase in motivation and interest for students who were taught with CAI. Results may be generalized to classrooms in urban settings where students consist primarily of minority populations.

In the research work “Enhancing Tenth Graders' Earth-Science Learning through Computer-Assisted Instruction”, Chang, Chun-Yen (2000) investigated the comparative efficiency of computer-assisted instruction (CAI) and traditional teaching methods on 10th grade students' learning of earth science in Taiwan.

Efficacy of CAI was compared with regular classroom teaching, studying from the textbook, or from plain text files on the computer in the study “Science Teaching through Computer Assisted Instruction: Research Findings and Insights” by Mridula D. Ranade (2001) in which Single group and control/comparison group pretest-post test designs were used. The response of students to CAI has been overwhelmingly positive. ‘t’ tests for comparison of pre and post-test means have revealed that CAI has in every case led to increased achievement.

To determine the retention effect of Computer Assisted Instruction (CAI) on students' academic achievement for teaching the Physics topics, Kara, Izzet (2008) studied the impact of teacher supervised CAI method on school students in comparison with Traditional instruction (TI) method (control) and found Significant differences
between the Science Subject test scores of experiment and control group were found in favor of experiment group after 5 months which clearly indicates the retention effect of Computer Assisted Instruction.

The effectiveness of the BBTA (brain-based teaching approach) in dealing with issues related to the learning motivation towards the subject of physics amongst secondary school students in Malaysia was studied and the study showed that the BBTA module was an effective teaching approach in dealing with the issue aforementioned. It was found that students who followed the BBTA module possessed a better physics learning motivation compared to students who received CTM (conventional teaching method) (Saleh, Salmiza, 2011).

Certain teaching dispositions might play a major role in improving student science achievement in urban schools. Moreover, the findings of this study might help to further inform and guide policymakers in the transformation of educational policies that affect schools and science educators in urban setting Miranda, Rommel, (2012),

2.2.3. CAI Effectiveness

In this sub section studies related to effectiveness of CAI on teaching and learning are discussed:

Under the title “Research on the Effectiveness of Computer-Based Instruction: A review”, Hasselbring, (1984) summarized results of research studies and Meta analyses on the effects of computer-based instruction on student achievement and attitudes. Results favored the use of CBI over traditional instruction.

Bracey, (1987) conducted a study on "Computer-Assisted Instruction: What the research Shows", and provided a brief summary of research conducted since the author's previous review on this topic, conducted in 1982. Main findings: 85-95 percent
of studies showed positive effects of CAI, and the effectiveness of CAI decreased from the elementary to secondary to post secondary level.

Roblyer, Castine and King (1988) in their work “Assessing the Impact of Computer-Based Instruction: A Review of Recent Research” described the methodology and findings from a meta analysis of 82 studies and dissertations on the use of microcomputers in education from the elementary through college and other adult levels. Research generally indicated favorable achievement effects, but with some notable exceptions, such as ESL.

Lowe, Nedra; Bickel, Robert (1993) examined whether computer-assisted instruction (CAI), compared to conventional instruction, would improve the teaching of grammar and writing to college students studying developmental communication. A quasi-experimental pretest/post test evaluation indicated students receiving CAI enjoyed a substantial advantage in the presence of a fairly broad range of control variables.

Cavanaugh (2001), in a “Quantitative Synthesis of Recent Research on the Effects of Teaching and Learning With Technology on Student Outcomes”, estimated the effects of teaching and learning with technology. This result indicated that teaching and learning with technology had a small, positive, significant ($p < .05$) effect on student outcomes when compared to traditional instruction. On the other hand, the study indicated that technology had a small, negative effect on students' behavioral outcomes.

Bayraktar and Sule (2002) conducted “A Meta-Analysis of the Effectiveness of Computer-Assisted Instruction in Science Education”. Results showed a small positive effect for CAI use when used in simulation or tutorial modes, with individual computer use, and when used as a supplement to traditional instruction.
Student responses to a tertiary, flexibly delivered physics course were examined and evaluations indicate that it has been highly successful overall, but also reveal something of how participants responded to different modes of delivery. Both students and lecturing staff valued highly the opportunity to experience a more traditional mode of teaching and learning provided through an intensive residential component of the course (Woolnough, Jim, et.al., 2006).

“Is the computer based physics instruction as effective as laboratory intensive physics instruction with regards to academic success on electric circuit 9th grade students?” At the end of the study, it was concluded that the computer based learning is as effective as the laboratory based learning on students’ achievement (Bekir Bayrak, et.al., 2007).

Celal Bayrak, (2008) investigated whether computer assisted instruction was more effective than face-to-face instruction in increasing student success in physics. Findings revealed that the experimental group which had the instruction through the computer simulation was more successful than the control group who had face-to-face instruction.

Boucheix, Jean-Michel & Schneider, Emmanuel (2009) investigated how learners comprehend the functioning of a three-pulley system from a presentation on a computer screen. The experiments indicated that an animation as well as integrated sequential static frames enhanced comprehension and showed that a controllable animation did not have a powerful effect on comprehension, except for learners with low spatial and mechanical reasoning abilities.

Amusement park physics is a popular way to reinforce physics concepts and to motivate physics learners. Research into the impact of participating in the competition
revealed positive effects such as the acquisition of experimentation skills and improved attitudes towards physics (Moll, Rachel; 2010).

Analysis of achievement tests showed a significant difference between the students achievements at the knowledge and comprehension levels of cognitive domain (P<0.05). On the other hand no difference was noted between their achievements at the application level of cognitive domain. The results have indicated that the student’s attitudes towards physics learning were not affected by different instruction methods (Selahattin Gonen, et.al., 2010)

Carr, David & Bossomaier, Terry (2011) described embedding the physics of relativity into a computer game, and presented the results of a study on its effectiveness for learning. It was found that the game on its own serves as a powerful introduction for building up accurate qualitative descriptions of relativistic physics effects; in addition, learners generally reported finding the game accessible and interesting. However, establishing deeper understanding of the physics requires further reflection on the part of the learners than the game itself tends to facilitate.

Guzel, Hatice (2011) studied the factors motivating secondary school physics teachers and found that the most motivating factor for physics teachers was factors related with advantages and the least motivating factor was related with the profession. The interesting and challenging thing here is that the factors "related with profession" were at the sixth and so the final stage. Since factors such as making different lectures, following new source materials, information exchange with colleagues, presence of education technology and growing good students were at the final stage for motivation of teachers, this presents serious challenges in terms of future of teaching profession and growing of rising generation.
Physics classes, due to the extensive amount of abstracts concepts, are at the top of the classes causing students difficulties. Therefore, computer supported instruction will facilitate the comprehension of the students which would otherwise be difficult to understand. A study by Guzel, Hatice (2011) showed that the majority of physics teachers believed that the use of computers facilitates learning.

The most significant findings by Brecht, H. David, (2012), are that video lectures are used by students for tutorial help, they improve initial learning and the weakest students are especially benefited, with significantly reduced course withdrawal rates.

2.2.4. CAI Meta Analysis

In this section the review articles related to CAI are discussed. Meta-analysis as a technique is an integrative statistical analysis or reanalysis of previous research as a means to answer new questions using old data (Glass et al., 1981). It combines results from multiple similar individual research studies to generate a single effect size that illustrates the treatment effect across all studies, to make studies comparable.

The Meta analysis by the North West Regional Educational Labs entitled “Does computer-assisted instruction have proven benefits? For example, does it result in higher learning gains?” summarized 59 CAI research reports for K-12 and provided the following research findings from (1985 -1990):

(i) The use of CAI as a supplement to conventional instruction produces higher achievement than the use of conventional instruction alone. (ii) Computer-based education (CAI and other computer applications) produce higher achievement than conventional instruction alone. Students learn material faster and retain what they have learned with CAI than with conventional instruction alone. (iii) The use of CAI leads to
more positive attitudes toward computers, course content, quality of instruction, school in general, and self-as-learner than the use of conventional instruction alone. (iv) The use of CAI is associated with other beneficial outcomes, including greater internal locus of control, school attendance, motivation/time-on-task, and student-student cooperation and collaboration than the use of conventional instruction alone. (v) CAI is more beneficial for younger students than older ones and CAI is more beneficial with lower-achieving students than with higher-achieving ones. (vi) Economically disadvantaged students benefit more from CAI than students from higher socioeconomic backgrounds. (vii) There are no significant differences in the effectiveness of CAI with male and female students. (viii) Students' fondness for CAI activities centers on the immediate, objective, and positive feedback provided by these activities. (ix) CAI activities appear to be at least as cost-effective as—and sometimes more cost-effective than other instructional methods, such as teacher-directed instruction and tutoring.

Kulik (1994) synthesized findings from 546 individual studies at all levels of CBI implementation conducted during the period 1978 - 1991. He conducted a meta-analysis of 12 CBI meta-analytic studies based on 546 individual studies and reported that students learned more in less time when they received computer-based instruction, and that students liked their classes more and developed more attitudes that are positive when their classes include computer-based instruction. These findings suggested that students who engaged in CBI performed significantly better than students who did not. However, CBI did not have positive effects in every area in which they were studied. CAI is more effective in improving achievement of younger students and students with special learning needs.
In an updated meta-analysis on the learning effect of CAI based on 120 individual studies, Fletcher-Flinn and Gravatt (1995) examined the effect of a range of variables that related to treatments, methodologies, educational level, course content, and student characteristics. Their results and estimates were similar to previous reviews showing a learning benefit for CAI suggesting that CAI students would outperform 60 percent of the students from traditional classes. However, no significant difference in effect size between educational level was found, though the effect for elementary level was slightly larger. This pattern of results is inconsistent with Kulik's model and those of other earlier reviews, the authors suggested that it may reflect a decline in novelty as computers become more commonplace in schools and at home.

Christmann, Badgett and Lucking (1997) compared the academic achievement of students in grades six through twelve who received either traditional instruction or traditional instruction supplemented with CAI across eight curricular areas. The comparative effectiveness of CAI showed that the effect size for science subject was the highest and English the lowest, whereas mathematics was in between the other two.

In summary, findings from both reviews and meta-analyses showed a general learning advantage for CAI over traditional instruction, though the level of effectiveness of CAI may vary with specific student population, course content and CAI type. Another explanation is that the advantage stems from the generally superior quality of CAI materials, rather than from some intrinsic characteristics of the computer technology as a vehicle of instruction (Fletcher-Flinn & Gravatt, 1995).

In an another review that did not employ meta-analysis, Sivin-Kachala (1998) assessed the effect of computer technology on learning and achievement by analyzing 219 individual research studies conducted from 1990 to 1997 across all learning domains and all learner ages. He reported that a) students in technology rich
environments experienced positive effects on achievement in all major subject areas; b) students in technology rich environments showed increased achievement in preschool through higher education for both regular and special needs children; and c) students' attitude toward learning and their own self-concept improved consistently when computers were used for instruction. However, he acknowledged that the level of effectiveness of educational technology is influenced by the specific student population, the software design, the educator's role, and the level of student access to the technology.

Bayraktar and Sule (2002) conducted “A Meta-Analysis of the Effectiveness of Computer-Assisted Instruction in Science Education.” This meta-analysis investigated how effective computer-assisted instruction (CAI) is on student achievement in secondary and college science education when compared to traditional instruction. Results showed a small positive effect for CAI use when used in simulation or tutorial modes, with individual computer use, and when used as a supplement to traditional instruction.

Pol, Henk., et. al. (2009) investigated the effectiveness of computer-delivered hints in relation to problem-solving abilities in two alternative indirect instruction schemes. The instruction schemes are supplied by means of a web-based program, which supports the development of strategic knowledge [Pol, Harskamp, & Suhre, (2008). "The effect of timing in instructional support, for computer supported problem-solving programs, for students in secondary schools." Results of the experiment show that both computerized instruction schemes are effective. Students working with the most elaborate instruction scheme show an increased use of their pallet of heuristics and algorithms in the post-test. Furthermore, the instruction scheme in which hints are
available to students during problem-solving proves to be most effective when students show an increase in the systematic use of hints during problem-solving.

2.3. RESEARCHES RELATED TO COMPUTER ADAPTIVE TESTING

Teacher educators need constant input of processes related to Quality improvement in their career like refresher courses, and special trainings in leadership and human resource development so that they can be vibrant sources to influence their wards effectively. Teaching as a profession and Teacher professional development through programmes are realised as the need of the hours when we claim that for human development modernization of education and Technology involvement in various processes in education are the felt needs says Madhumathi (2006). Computer technology can open up opportunities for developing innovative assessment tools in science education. The nature of computers as information processing tools, the role of computer technology in 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.

This section deals with the summary of the literature related to the Computer Adaptive tests (CAT) under the following headings:

- Computerized Adaptive Test Development
- Procedures of CAT, IRT
- Test Designs in CAT
- Effectiveness of CAT
- Gender Difference in CAT/ CAI
- Students’ Attitudes towards CAT
- Learning Time
- Learning Rate

2.3.1. Computerized Adaptive Test Development

In this section studies related to different methods of developing Computer Adaptive Test Packages are discussed:
Bergstrom, Betty; Stahl, John (1992) reported in their work “Assessing Existing Item Bank Depth for Computer Adaptive Testing” a method for assessing the adequacy of existing item banks for computer adaptive testing. An example was presented which showed that the adequacy of the bank could depend on the stopping rule implemented. The procedure could be modified for use with other Item Response Theory models as long as the item parameters are known.

Carlson, Randal D. (1994) reviewed the development and use of mass testing procedures for evaluating large groups of relatively homogeneous individuals and suggested computer adaptive testing as an alternative to conventional testing. Highlights included the evaluation of public education and higher education; needs assessment; summative evaluation; adaptive testing; and future possibilities in their work “Computer Adaptive Testing: A Shift in the Evaluation Paradigm.”


“Some Considerations for Eliminating Biases in Ability Estimation in Computerized Adaptive Testing”. - Work done by Samejima, Fumiko (2000). In this work, Item response theory (IRT) had been adapted as the theoretical foundation of computerized adaptive testing (CAT) for several decades. The essential issues of CAT were addressed in this paper, and then several ways of eliminating noise and bias in estimating the individual parameter, theta, of person “a” are proposed and discussed, so that accuracy and efficiency in ability estimation could be increased. It was suggested
that several graded response items be used at the beginning of the CAT to avoid the influence of bias and lack of information inherent in dichotomous response items.

Dunkel, Patricia. (2000) worked on the title “Considerations in Developing and Using Computer Adaptive Tests to Assess Second Language Proficiency” and reported that, in a CAT, each examinee took a unique test that was tailored to his or her ability level. Avoided were questions that had low information value about the test taker’s proficiency. The result of this approach was higher precision across a wider range of ability levels.

“Multi method Analysis of Mathematics Achievement Tests” report prepared by Dimitrov, Dimiter (2000) combined Item Response Theory (IRT) and statistical methods. It provided information to educational analysts about whom IRT model fits.

“Integrating Stratification and Information Approaches for Multiple Constrained CAT” was studied by Leung, Chi-Keung; Chang, Hua-Hua; Hau, Kit-Tai (2001). This study believed that item selection methods using the maximum information approach (MI) could maintain high efficiency in trait estimation by repeatedly choosing high discriminating (alpha) items. The performances of MI, BASTR, and their integration, MIBASTR, were empirically compared through simulation. Results indicated that BASTR was the best in utilizing the entire pool evenly, and thus in tackling item security problems. On the other hand, MI and MIBASTR offered high and comparable measurement efficiency. The latter outperformed the former in item exposure and pool utilization.

Papanastasiou & Elena C. (2002) presented a report on A “Rearrangement Procedure” for Scoring Adaptive Tests with Review Options. This study tested the efficacy of a rearrangement procedure that rearranged and skipped certain items in order to estimate the examinees abilities better, without allowing them to cheat on the
test. Results showed that the rearrangement procedure was effective in reducing the bias of the ability estimates. The reliability decreased slightly after the procedure, but this decrease was negligible.

“Computerized Adaptive Testing: Some Issues in Development” - Orcutt, Venetia (2002) investigated this title and explained, the emergence of enhanced capabilities in computer technology. This paper provided a brief overview of how a CAT is developed, basic concepts of IRT and proficiency estimation, and examined a few of the issues associated with development of selected methodologies.

Wang, Xiang Bo; Pan, Weiqin; Harris, Vincent (2003) in their work, “Computerized Adaptive Testing Simulations Using Real Test Taker Responses”. It was hoped that through control of item parameters and examinee abilities based on a realistic test, insights would be gained on practical issues involved The 127 items were found to be sufficient to conduct the CAT sessions for virtually all test takers at the 3 accuracy levels. This study represented the first CAT simulation that used real test taker responses.

“An Evaluation of a Two-Stage Testlet Design for Computerized Testing. Law School Admission Council Computerized Testing.” – This report by Reese, Lynda; Schnipke, Deborah. (2003) explained a two-stage design provides a way of roughly adapting item difficulty to test-taker ability. The results indicated that if the test lets were carefully assembled, a two-stage test let design could produce more precise ability estimates in the middle ability range than those obtained from a paper-and-pencil design with twice as many items. Results of this study provided a baseline against which future research that incorporated content constraints could be compared.

Ronald Armstrong & Jennifer Edmonds(2004) investigated on “A Study of Multiple Stage Adaptive Test Designs” This work evaluated several multiple stage
adaptive test (MST) designs under the criteria of (a) accuracy of test scores, (b) simplicity of the design to facilitate review, and (c) efficiency of item pool usage to reduce the cost of item development. Results with an operational item pool were tabulated.

Computer-based testing provides an optimal format for creating individually-tailored tests. Ketterlin-Geller & Leanne, (2005) presents the steps involved in conceptualizing, constructing, and implementing a universally designed test in anticipation that test developers, state department assessment coordinators will benefit from this application with recommendations for future research and development efforts to create accessible computer-based learning environments for all students.

Kocakaya, Serhat & Gonen, Selahattin (2010) analysed and compared the physics questions of the university entrance exam (OSS) with those asked at exams at different schools in Turkey in terms of the levels of cognitive domain of Bloom's Taxonomy. In four types of high schools, it was found that 72.5 per cent of the questions were of the lower-order cognitive skills (LOCS) type. On the other hand, about half of the questions asked in the university entrance examination (OSS) were of the higher-order cognitive skills (HOCS) type. This contradiction causes a problem between the assessment at high school and that at the OSS.

2.3.2. Procedures of CAT/ IRT

In this section studies related to the procedures, components needed for developing computer adaptive tests are listed out: In a study with the title “Test-Retest Consistency of Computer Adaptive Tests” Lunz, Mary. et al. (1990) explored the test-retest consistency of computer adaptive tests of varying lengths. The ability measures from the test and retest were found to correlate at 0.95 when attenuated for error, demonstrating that differentiation among
examinee measures was comparable regardless of the length of the test or the particular subset of items. This finding provided evidence of test-retest consistency of computer adaptive tests.

In their research work “The Effect of Item Pool Restriction on the Precision of Ability Measurement for a Rasch-Based CAT: Comparisons to Traditional Fixed Length Examinations” Halkitis, Perry. (1998) reported that regardless of the item pool size, CAT provides greater precision in measurement with a smaller number of items administered even when the choice of items was limited, but CAT failed to achieve equiprecision along the entire ability continuum.

In a study on “The Effects of Judgment-Based Stratum Classifications on the Efficiency of Stratum Scored CATs” Finney, Sara; Smith, Russell; Wise, Steven. (1999) it was found that stratum CATs based on empirical item difficulties (both p-values and b-parameters) had increased efficiency and precision relative to a conventional fixed-length test. It was also shown that efficiency and precision increased as the number of strata increased. Also, under certain conditions, the stratum CAT was able to match or exceed the precision and efficiency of the traditional CAT. Findings suggested that stratum CATs based on human judgments do not provide increased efficiency or precision over a conventional fixed-length test.

Van Der Linden and Wim. (2000) investigated on “Optimal Stratification of Item Pools in a-Stratified Computerized Adaptive Testing” and presented a method based on 0-1 linear programming (LP) to stratify an item pool optimally for use in "alpha"-stratified adaptive testing. The results indicated that the new method performed well in practical situations. It improved item exposure control, reduced the mean squared error in the theta estimates, and increased test reliability.
In their study “Content Balancing in Stratified Computerized Adaptive Testing Designs” Leung, Chi-Keung; Chang, Hua-Hua; Hau and Kit-Tai (2000) studied the item selection methods in computerized adaptive testing (CAT). Chang and Ying (1999) proposed the a-stratified design (ASTR) that attempts to equalize item exposure distribution by uplifting the usage of low "a" items. The method had been demonstrated to be effective in improving the use of the entire pool, without sacrificing efficiency in ability estimation when it was used with certain types of item pools.

Bowles, Ryan; Pommerich, Mary (2001) investigated on the title “An Examination of Item Review on a CAT Using the Specific Information Item Selection Algorithm”. Many arguments had been made against allowing examinees to review and change their answers after completing a computer adaptive test (CAT). These arguments include: (1) increased bias; (2) decreased precision; and (3) susceptibility of test-taking strategies. Results of simulations suggested that the strength of these arguments was reduced or eliminated by using specific information item selection (SIIS).

Papanastasiou, Elena (2002) tested the efficacy of a rearrangement procedure. Results show that the rearrangement procedure is effective in reducing the bias of the ability estimates. The reliability decreased slightly after the procedure, but this decrease was negligible.

Pommerich, Mary (2002) addressing three questions: (1) Do examinees respond to items in the same way across administration modes and computer interface variations? (2) What are some of the factors that can contribute to modal effects? and (3) Can item parameters calibrated from paper and pencil administrations be used for computer administrations? in her study and found that while performance effects do occur across modes, they have a fairly small effect in practice. Simulation results
suggest that item parameters calibrated from paper and pencil tests could probably be used initially in a computer administration.

Rizavi, Saba et.al. (2002) investigated and quantified the tolerable error in item parameter estimates for different sets of items used in a computer-based testing. Results indicated that context effects played a more significant role in adaptive item parameters when comparisons were made to the parameters that were obtained from paper-and-pencil testing. This suggests that, whenever feasible, the parameter estimates obtained from paper-and-pencil administrations be replaced with computer-based testing calibrated parameters.

“Optimum Number of Strata in the a-Stratified Computerized Adaptive Testing Design” was studied by Hau, Kit-Tai; Wen, Jian-Bing; Chang and Hua-Hua (2002). In the a-stratified method, a popular and efficient item exposure control strategy proposed by Chang and Ying, (1999); Hau and Chang, (2001) for computerized adaptive testing (CAT), the item pool and item selection process had usually been divided into four strata and the corresponding four stages. In a series of simulation studies, researchers examined the optimum number of strata. Results showed that quite independent of the item pool size and the correlation between item discrimination and difficulty, ability estimation deteriorated while the number of over- and under-exposed items decreased with an increase in stratum number. There was a diminishing return in that dividing the pool into too many strata could also be problematic because when the stratum was too small, there were not any items of close difficulty for each particular examinee. The results were in general agreement with the speculation that too few and too many strata may not provide optimum efficiency and balanced item pool utilization. It was shown that the ideal and optimum number of strata to be used in each specific application depended on the item pool structure, test length, and other testing conditions.
In a study with the title “Calibrating CAT Pools and Online Pretest Items Using Marginal Maximum Likelihood Methods” Pommerich, Mary; Segall, Daniel (2003) conducted research to evaluate methods of calibrating pretest items for computerized adaptive testing (CAT) pools. In this study, pretest and operational CAT items were simultaneously calibrated and placed on the scale of the operational parameters from one CAT pool that was designated as the anchor CAT pool.

“Assessing the Efficiency of Item Selection in Computerized Adaptive Testing” was done by Weissman, Alexander (2003). They investigated the efficiency of item selection in a computerized adaptive test (CAT), where efficiency was defined in terms of the accumulated test information at an examinee's true ability level. The greatest increase in test efficiency was observed when the alternative ability estimation procedures (ML/Alt, MAP/Alt, and GSS) were used. The gains in efficiency were most pronounced for shorter tests, but were noticeable even for longer tests. Overall, it appeared that the ability estimation procedure had an impact on the efficiency of item selection to a large extent than the item selection procedure. Scores from CAT tests are to be used for a variety of purposes, from identifying whether individual students are proficient, to helping determine whether schools are causing adequate growth for their students (Kingsbury, Gage & Hauser, Carl., 2004).

The study by Tian, Jian-quan et.al, 2007 stressed that the Computerized adaptive testing (CAT) has unsurpassable advantages over traditional testing, which has become the mainstream in large scale examinations in modern society and gives a brief introduction to CAT including differences between traditional testing and CAT, the principles of CAT, psychometric theory and computer algorithms of CAT, the advantages and cautions of CAT. It also reviews the development of CAT in China.
Since researchers acknowledged the several advantages of computerized adaptive testing (CAT) over traditional linear test administration, the issue of item exposure control has received increased attention. Due to CAT's underlying philosophy, particular items in the itempool may be presented too often and become overexposed (Georgiadou, Elissavet et.al, 2007).

The study by Veldkamp, Bernard. et.al (2010) on A Multiple Objective Test Assembly Approach for Exposure Control Problems in Computerized Adaptive Testing revealed that Information in the test has to be maximized, item compromise has to be minimized, and pool usage has to be optimized to avoid the problems such as Overexposure and underexposure of items in the bank in operational computerized adaptive testing (CAT) systems.

Automated feedback for objective testing system has been further developed for marking practical and essay questions and providing automated feedback. Recent research at the University of Hertfordshire was able to show that learners and tutors accept and value our automated feedback approach based on objective tests and Computer Adaptive Testing. The provision of fast effective feedback is vital and this system was found to be an important addition to the tools available Barker &Trevor (2011).

IRT STUDIES

Monitoring item behavior in a timely manner is extremely important in CAT for practitioners to take appropriate actions, such as blocking problematic items from active use or pulling the items from subsequent item pools (Zhu, Renbang et.al, 2002).

The examination of the influence of CAT data on item response theory three-parameter logistic model (3PL) item characteristic curves and on item parameter estimates for the 3PL model in a Bayesian context showed that the error in the item
characteristic curve as a whole and in the estimates of item difficulty appeared to be reduced when CAT data information was incorporated into the estimation of the parameter suggesting that error in estimating item characteristic curves may be reduced by incorporating CAT information into the calibration process (Smith, Robert et.al, 2002).

Preventing items in adaptive testing from being over- or underexposed is one of the main problems in computerized adaptive testing. Using a system of rotating item pools, on the other hand, is a method that potentially solves both problems. Ariel, Adelaide et.al (2002) developed a test assembly model for the problem of dividing a master pool into a set of smaller pools. The model was motivated by Gullicksen's (1950) matched random subtests method. Different methods to solve the model are proposed. An item pool from the Law School Admission Test was used to evaluate the performances of computerized adaptive tests from systems of rotating item pools constructed using these methods.

A Comparison of Exposure Control Procedures in CAT Systems Based on Different Measurement Models for Testlets Using the Verbal Reasoning Section of the MCAT. Worked by Boyd, Aimee.et.al; (2003). In this study compared several item exposure control procedures for computerized adaptive test (CAT) systems based on a three-parameter logistic testlet response theory.

CATSIB: A Modified SIBTEST Procedure to Detect Differential Item Functioning in Computerized Adaptive Tests. Law School Admission Council Computerized Testing Report prepared by Nandakumar, Ratna; Roussos, Louis (2003). This paper proposed a modification of the SIBTEST DIF procedure for CATs, called CATSIB. A simulation study was conducted to evaluate the performance of CATSIB. Results showed that CATSIB with the regression correction displayed impact-induced
Type I error inflation. In terms of power, even with as few as 250 test takers in each group, CATSIB had detection rates of 64% or greater for large values of DIF. When sample size was increased to 500 in each group, these power rates increased to more than 90%. CATSIB displayed nearly unbiased estimation under nearly all the simulated conditions.

Impact of Local Item Dependence on Item Response Theory Scoring in CAT. Law School Admission Council Computerized Testing Report explained by Reese, Lynda M. (2003). This study represented a first attempt to evaluate the impact of local item dependence (LID) for Item Response Theory (IRT) scoring in computerized adaptive testing (CAT). Results indicated that types of scoring an extreme amount of LID might adversely affect the final score obtained by the examinee. The estimated precision of the test was also affected by the extreme LID level studied here.

Reese, Lynda; Schnipke, Deborah (2003) worked “An Evaluation of a Two-Stage Testlet Design for Computerized Testing”. A two-stage design provided a way of roughly adapting item difficulty to test-taker ability. The results indicated that if the testlets are carefully assembled, a two-stage testlet design could produce more precise ability estimates in the middle ability range than those obtained from a paper-and-pencil design with twice as many items. Results of this study provided a baseline against which future research that incorporates content constraints can be compared.

“A Comparison of Testlet - Based Test Designs for Computerized Adaptive Testing”-prepared by Schnipke, Deborah; Reese, Lynda (2003). This study incorporated testlets (bundles of items) into two-stage and multistage designs, and compared the precision of the ability. Results with 50,000 and 25,000 simulated test takers indicated that all testlet-based designs resulted in improved precision over the same length paper-and-pencil test, and almost as much precision as the paper-and-
pencil test of double length. Given the many other (non psychometric) advantages of these designs, they might be viable options for computer-administered tests.

Variable-length Computerized Adaptive Testing (CAT) can provide examinees with tailored test lengths. The Study by Huo, Yan, (2009), explores four item selection methods under the variable-length situations and the results indicate that these four methods can achieve good ability estimation while maintaining balanced item usage in the variable-length CAT simulations. These study findings will advance the research and understanding of variable-length CAT, and will facilitate the application and adoption of variable-length CAT in real world testing.

Learning to read requires knowledge of word meanings for those words most commonly encountered in basic reading materials. Two randomized studies were conducted by Fehr, Charles Norman, (2011), to test the effects of an online, computer-adaptive vocabulary instruction program designed to provide remedial instruction on word meanings for high frequency words. Studies suggest that computer-delivered vocabulary instruction may be an efficient mechanism for remediation of vocabulary deficits.

To enhance the robust performance of CAT against aberrant responses, Barton and Lord proposed the four-parameter logistic (4PL) item response theory (IRT) model. However, studies conducted by Yen Yung-Chin; Ho, Rong-Guey; Laio, Wen-Wei; Chen, Li-Ju; Kuo, Ching-Chin, (2012), attempts to investigate the performance of the 4PL IRT model as a slip-correction mechanism with an empirical experiment. The results showed that the 4PL IRT model could not only reduce the problematic underestimation of the examinees’ ability introduced by careless mistakes in practical situations but also improve measurement efficiency.
An ordering-theory-based knowledge-structure-adaptive testing system was developed and evaluated by Wu, Huey-Min; Kuo, Bor-Chen; Yang, Jinn-Min, (2012). The results of this system showed that the two different interfaces, paper-based and computer-based, did not affect the examinees' performance. In addition, the effect of correct guessing was discussed, and two methods with adaptive testing algorithms were proposed to mitigate this effect. The experimental results showed that the proposed methods improve the effect of correct guessing.

2.3.3. Test Designs in CAT

In this section various design of CAT are discussed:

In a research work with the title “Learner Control versus Computer Control in Instructional Simulation” Mattoon, Joseph et.al (1991) reported an experiment designed to assess the effects of learner control over the level of challenge in computer-assisted instruction (CAI) and collect data that would guide the design of future learner control experiments. The subjects were 78 undergraduate students in the College of Education at a large public university. The CAI included a simulation of an aviation instrument and taught subjects how to locate an aircraft's position in space by interpreting a complex display. Instructional research had not yielded the empirical information necessary to develop prescriptions for employing learner control features in this type of CAI. Three treatments that varied the subject's control over the level of challenge were compared: learner control, computer control, and learner control with advisement. Results indicated that learner control did affect subjects' time on task on an immediate test, but this effect appeared to be temporary. The change in time on task was probably caused by the way subjects adjusted their challenge level. Subjects who received advisement tended to set challenge levels higher and adjusted the challenge more frequently.
Bergstrom, Betty and others (1992) undertook a work entitled “Altering the Level of Difficulty in Computer Adaptive Testing” and studied the effects of altering test difficulty on examinee ability measures. The test length in a computer adaptive test was studied for 225 medical technology students in 3 test difficulty conditions. Results suggested that, with an item pool of sufficient depth and breadth, acceptable targeting to test difficulty was possible.

Weiss and Kingsbury (1984) and more recently Spray and Reckase (1994) described CATs for situations where the main interest was not in estimating the ability of an examinee, but to classify the examinee in of two categories, e.g., pass-fail, master/non-master. The main results of the study were that a gain of at least 25% in the mean number of items was to be expected in a CAT. Furthermore, it was concluded that for the three way classification problem using statistical testing was the most promising computation procedure. Finally, it was concluded in this case that imposing the item selection with constraints in the form of content and/or exposure control hardly impaired the quality of the testing.

In a work on “The Accuracy of Examinee Judgments of Relative Item Difficulty: Implications for Computerized Adaptive Testing” Wise, Steven and others (1997) experimented the degree to which item review on a computerized adaptive test (CAT) could be used by examinees to inflate their scores artificially. G. G. Kingsbury (1996) described a strategy in which examinees could use the changes in item difficulty during a CAT to determine which items' answers were incorrect and should be changed during item review. The results of two studies involving groups of 77 and 62 undergraduates suggested that examinees were not highly proficient at discriminating item difficulty, a skill needed for successful application of the Kingsbury strategy. In the third study, which used 243 introductory statistics students, the Kingsbury strategy,
which examinees would use only for guessed items, was compared to a generalized strategy used for all sequential item pairs. The Kingsbury strategy yielded a small average score gain, while the generalized strategy yielded an average score loss. These results suggested that only the Kingsbury strategy would enable examinees to inflate their scores successfully.

The study entitled “A Comparison of Testlet-Based Test Designs for Computerized Adaptive Testing” was conducted by Schnipke, Deborah; Reese, Lynda (1997). They found that two-stage and multistage test designs provided a way of roughly adapting item difficulty to test-taker ability. This study incorporated testlets (bundles of items) into two-stage and multistage designs, and compared the precision of the ability estimates derived from these designs with those derived from a standard CAT design and from paper-and-pencil test designs. Results indicated that all testlet-based designs resulted in improved precision over the same-length paper-and-pencil test, and almost as much precision as the paper-and-pencil test of double length. Given the many other (non psychometric) advantages of these designs, they might be viable options for computer-administered tests.

In his research work “The Relative Efficiency of Two-Stage Testing Versus Traditional Multiple Choice Testing Using Item Response Theory in Licensure” Reed . Castle (1997) applied two-stage testing (TST) to a licensure testing examination. The results of this study indicated that the 3p-F model was associated with decision consistency indices superior to the 3p-V model. Increased items in either the Routing or Measurement test, led to higher kappa values. Although actual routing errors were high (13% to 31%), IRT appeared to rectify some of the negative effects of routing error in terms of decision consistency.
In their study “Incorporating Content Constraints into a Multi-Stage Adaptive Testlet Design”, Reese, Lynda; Schnipke, Deborah; Luebke, Stephen (1999) realized that most large-scale testing programs facing computerized adaptive testing (CAT) must face the challenge of maintaining extensive content requirements, but content constraints in computerized adaptive testing (CAT) could compromise the precision and efficiency that could be achieved by a pure maximum information adaptive testing algorithm. The results revealed that constraints to control for item exposure, testlet overlap, and efficient pool utilization need to be incorporated into the testlet assembly algorithm. More refinement of the statistical constraints for testlet assembly was also necessary. However, even for this preliminary attempt at assembling content-balanced testlets, the two-stage computerized test simulated with these testlets performed quite well.

“An Evaluation of a Two-Stage Testlet Design for Computerized Testing” was undertaken by Reese, Lynda; Schnipke, Deborah (1999) to explore a two-stage design that provided a way of roughly adapting item difficulty to test-taker ability. The results indicated that if the testlets are carefully assembled, a two-stage testlet design can produce more precise ability estimates in the middle ability range than those obtained from a paper-and-pencil design with twice as many items. Results of this study provided a baseline against which future research that incorporates content constraints could be compared.

By using the work on “Improvement of Instruction with the Use of CATs To Clarify Student Learning Strategies” Hoegl, Juergen (1999) improved instruction in a community college course on Latin American culture and civilization, an instructor designed three interrelated classroom assessment techniques (CATs) to study student reactions to instructional approaches and student learning strategies. Results from the
first CAT validated as useful and effective the instructor's teaching approach through multiple methods to engage students in analyzing and reflecting about course readings. Taken together, responses to the three tests seem to identify a gap in assumptions about student learning strategies (at least for reading course texts) in which the instructor might assume that students engage the text with specific learning techniques, while the students were not actually using these techniques and might be expecting the instructor to initiate or lead them through the learning process.

In the study “Adaptation of a-Stratified Method in Variable Length Computerized Adaptive Testing” Wen, Jian-Bing; Chang, Hua-Hua; Hau, Kit-Tai (2000) felt that test security had often been a problem in computerized adaptive testing (CAT) because the traditional wisdom of item selection overly exposed high discrimination items.

Froelich, Amy Goodwin (2000) developed a “Computerized Adaptive Testing for Classifying Examinees into Three Categories”. The number of applications of computerized adaptive testing based on Item Response Theory (IRT) was growing quickly and psychometric research on adaptive testing was getting widespread attention. Traditionally a computerized adaptive test (CAT) aims at the efficient estimation of an examinee’s ability. However, it also had shown to be a useful approach to classification problems.

The a-stratified (STR) design advocated by Chang (2002) and his collaborators, which used items of less discrimination in earlier stages of testing, had been shown to be very successful in balancing and hence maximizing the usage of all items in the pool. In this series of simulation studies with variable length CAT in which testing terminated at a targeted test information, researchers examined whether the use of more items in STR to attain accuracy similar to the maximum information (Max-I) approach
in ability estimation would result in a greater exposure of all items. Simulations with self-generated items as well as an operational pool supported the usefulness of the STR method in general. However, the results suggested that it was desirable to have fewer in number but less discriminating items at earlier stages of testing and had a larger number of highly discriminating items at later stages.

“Mathematical-Programming Approaches to Test Item Pool Design” by Veldkamp, Bernard; Vander Linden; Ariel, Adelaide (2002) presented an approach to item pool design that had the potential to improve on the quality of current item pools in educational and psychological testing and thus to increase both measurement precision and validity. The paper concluded with an empirical example of the problem of designing a system or rotating item pools for CAT.

Spray, Judith et.al (2002) proposes an alternative method for controlling item allocation that is based on randomization. An example from an actual item pool is presented to illustrate the method. Results show that it is possible to control the overall allocation of items across multiple test forms assembled through automated assembly methods using the same procedure that is used to control for item exposure in computerized adaptive testing situations. The iterative procedure was programmed directly into the form-assembly code, so that iterations become part of the assembly process. The goal is to produce the desired item allocation across forms, rather than to obtain exposure-control parameters for each item.

The bundled multistage adaptive testing (BMAT) framework is a modification of the computer-adaptive sequential testing framework introduced by Luecht and. Nungester (1998). Under BMAT, banks of parallel testlets are constructed to meet various statistical targets and categorical constraints. BMAT requires automated test assembly (ATA) technology capable of handling multiple simultaneous objective
functions and constraint systems. The result is a secure method of building high-quality adaptive and mastery tests that have sever constraints on test content. BMAT is demonstrated in the context of high-stakes professional certification (Luecht, Richard, 2003).

The robustness of CAST when tests were constructed, administered, and scored by a uni-dimensional IRT model but item responses were multidimensional was evaluated. Results showed that, when multidimensionality was mild as measured by the angle distance between item clusters, uni-dimensional ability estimates and routing decisions were not sensitive to the level of content misclassification in item pools. (Zhang, Yanwei & Nandakumar, Ratna, 2006).

In order to make learning environments more efficient, Wauters; Desmet; Van den Noortgate, (2010), have been exploring the possibility of an automatic adaptation of the learning environment to the learner or the context. One of the possible adaptation techniques is adaptive item sequencing by matching the difficulty of the items to the learner's knowledge level. This is already accomplished to a certain extent in adaptive testing environments, where the test is tailored to the person's ability level by means of the item response theory (IRT).

The feasibility of applying IRT in adaptive item-based ITSs was studied and discussed by Chatzopoulou, D. I.; Economides, A. A., (2010). The two main challenges associated with IRT application in such learning environments: the challenge of the data set and the challenge of the algorithm. It was concluded that applying IRT seems to be a viable solution for adaptive item selection in item-based ITSs provided that some modifications are implemented. This paper presents Programming Adaptive Testing (PAT), a Web-based adaptive testing system for assessing students' programming knowledge. PAT classifies a student in one out of three programming
skills' levels. It can predict the corresponding classification of students in Greek National Exams. Furthermore, it can be helpful to both students and teachers. A student could discover his or her programming shortcomings. Similarly, a teacher could objectively assess his or her students as well as discover the subjects that need to be repeated.

Differential item functioning (DIF) occurs when an item on a test, questionnaire, or interview has different measurement properties for one group of people versus another, irrespective of true group-mean differences on the constructs being measured. item response theory based likelihood ratio testing for DIF (IRT-LR or IRT-LR-DIF). A simulation study using new variation, called Ramsay-curve DIF (RC-DIF), was conducted which showed that these methods improve results compared with assuming normality mistakenly, and an empirical example illustrates the procedures. In RC-DIF, the latent density for both groups is estimated simultaneously with the item parameters (Woods, Carol , 2011).

Yen, Yung-Chin; Ho, Rong-Guey; Laio, Wen-Wei; Chen, Li-Ju; Kuo, Ching-Chin, (2012), research findings showed that the 4PL IRT model could not only reduce the problematic underestimation of the examinees' ability introduced by careless mistakes in practical situations but also improve measurement efficiency.

2.3.4. Effectiveness of CAT

In this section studies related to the effectiveness of using computers as an evaluative tool, which improves learning are listed out:

Collins, Michael (1984) in their study “Improving Learning With Computerized Tests” compared Instructor's opinion, student questionnaires, differences in mean test scores of computer and non computer classes, and differences in performance on written class tests between users and nonusers of computer tests were analyzed to
investigate whether learning improved with computerized tests. Suggestions on how teachers might make use of computerized tests were included.

“Applying Computerized Adaptive Testing in Schools” by Olsen, James (1990) presented two studies applying computerized adaptive testing (CAT) in schools. Compared paper-administered, computer-administered, and CAT modes for administering school achievement and assessment tests. Then compared computerized adaptive aptitude test results with individually administered Weschler Intelligence Scale for Children-Revised. Found that computerized adaptive tests had comparable results, in less time, with high correlations to paper tests.

In their research work with the title: “Computerized Adaptive Measurements of Attitudes” Koch, William and others (1990) implemented computerized adaptive testing (CAT) to measure students' attitudes toward alcohol. They administered a paper-and-pencil version and a CAT version of attitudes towards alcohol scale to 113 undergraduates enrolled in health education classes. Findings showed a high correlation between scores from the CAT and the paper-and-pencil versions.

The study on “Confidence in Pass/Fail Decisions for Computer Adaptive and Paper and Pencil Examinations” by Bergstrom, Betty; Lunz, Mary (1991) examined the level of confidence in pass/fail decisions obtained with computer adaptive tests (CATs) was compared to decisions based on paper-and-pencil tests. The tests used in this study constituted part of the subjects' review for the certification examination The CAT was stopped when the examinee measure was 1.3 times the error of measure above or below the pass/fail point (one tailed 90% confidence interval) or when a maximum test length was reached. Results showed that greater confidence in the accuracy of pass/fail decisions was obtained for more examinees when the CAT implements a 90% confidence-stopping rule than with paper-and-pencil tests of comparable length. Using
a confidence interval stopping rule with a CAT allowed each examinee to take a minimum number of items and insured that pass/fail decisions were made with a high level of confidence.

Parshall, Cynthia ; Kromrey, Jeffrey. (1993) conducted a research on “Computer Testing versus Paper-and-Pencil Testing: An Analysis of Examinee Characteristics Associated with Mode Effect.” This paper analysed whether examinee characteristics were systematically related to mode effect across paper and computer versions of the same instrument, using data from the Graduate Record Examination (GRE) of the Educational Testing Service in its Computer-Based Testing Pilot Study of 1991. The following characteristics of 1,114 examinees were studied. For this method, mode effect was treated as a three-level, categorical, independent variable. Data demonstrate mode effect and support the conception of a small subset of examinees whose performance was more affected by mode than that of the total sample. The search for examinee characteristics that explained occurrence of mode effect, however, yielded inconsistent results, with only weak relationships to mode effect.

“Linking the Standard and Advanced Forms of the Raven's Progressive Matrices in both the Pencil-and-Paper and Computer-Adaptive-Testing Formats” was studied by Styles, Irene; Andrich, David (1993). This paper described the use of the Rasch model to help implement computerized administration of the standard and advanced forms of Raven's Progressive Matrices (RPM), to compare relative item difficulties, and to convert scores between the standard and advanced forms. The sample consisted of 95 girls and 95 boys in Australia.

In a study with title “An Empirical Study of Computerized Adaptive Test Administration Conditions” Lunz, Mary ; Bergstrom, Betty (1994) studied the impact of computerized adaptive test (CAT) administration formats on student performance
with 645 medical technology students who also took a paper-and-pencil test. Analysis of covariance indicated no significant interactions among test administration formats and provided evidence for adjusting CAT test to more familiar modalities.

Under the title “Test Anxiety and Test Performance: Comparing Paper-based and Computer Adaptive Versions of the Graduate Record Examinations (GRE) General Test” Powers, Donald E. (2001) tested the hypothesis that the introduction of computer-adaptive testing might help to alleviate test anxiety and diminished the relationship between test anxiety and test performance. They compared a sample of Graduate Record Examinations (GRE) General Test takers who took the computer-adaptive version of the test with another sample who took the paper-based version. There was no support for the study's major hypothesis.

A CAT will result in examinees answering different number of items according to the stopping rule employed. Unfortunately, the realities of testing conditions such as scheduling and improper test-taking strategies on the part of examinees have necessitated the imposition of time and minimum test length limits on CATs. Such constraints might place a burden on the CAT test taker resulting in aberrant response behaviors by some examinees. Occurrence of such response patterns result in inaccurate estimation of examinee proficiency levels. Rizavi & Hariharan (2001) examined the effects of test length, time limits and their interaction with the examinee proficiency levels on the occurrence of aberrant response patterns due to rushed guessing.

In a computerized adaptive test (CAT), it would be desirable to obtain an acceptable precision of the proficiency level estimate using an optimal number of items. Decreasing the number of items is accompanied, however, by a certain degree of bias when the true proficiency level differs significantly from the a priori estimate.
Raiche (2000) has suggested that it is possible to reduce the bias, and even the standard error of the estimate, by applying to each provisional estimation one of a combination of these strategies: (1) the adaptive correction for bias proposed by Bock and Mislevy (1982); (2) adaptive a priori estimate; and (3) adaptive integration interval. A simulation study was conducted to explore these approaches. One thousand administrations of a CAT were simulated for varying proficiency levels. Expected a priori estimation of a unidimensional Rasch model with 40 quadrature points was used for provisional and final estimates. Simulation results demonstrate that bias can be reduced more effectively by using an adaptive adjustment of the estimation procedure than by considering only the final correction for bias proposed by Bock and Mislevy (Raiche, Gilles & Blais, Jean-Guy, 2002).

In a research paper with the title: “Performance Differences According to Test Mode and Computer Familiarity on a Practice Graduate Record Exam”, Goldberg, Amie; Pedulla, Joseph (2002) studied the relationship between test mode (paper and pencil or computerized with and without editorial control) and computer familiarity for 222 undergraduates. Results emphasized the importance of evaluating time constraints when converting exams from paper to computer delivery.

In the study “Paper-Based versus Computer-Based Assessment: Key Factors Associated with the Test Mode Effect”, Clariana, Roy; Wallace, Patricia (2002) described a study that seeks to confirm several key factors in computer-based versus paper-based assessment. Based on earlier research, the factors considered in this study of undergraduates, includes content familiarity; computer familiarity; competitiveness; and gender. Reports results of analysis of variance that showed the computer-based test group outperformed the paper-based test group.
Weissman, Alexander, 2003, investigated the efficiency of item selection in a computerized adaptive test (CAT), where efficiency was defined in terms of the accumulated test information at an examinee's true ability level. A simulation methodology compared the efficiency of 2 item selection procedures with 5 ability estimation procedures for CATs and observed that the greatest increase in test efficiency was when the alternative ability estimation procedures (ML/Alt, MAP/Alt, and GSS) were used. The gains in efficiency were most pronounced for shorter tests, but were noticeable even for longer tests. Overall, it appears that the ability estimation procedure impacts the efficiency of item selection to a large extent than the item selection procedure.

Ricardo Conejo, et.al. (2004) developed a system called SIETTE, and its theoretical bases are Computer Adaptive Testing and Item Response Theory. With SIETTE, teachers worldwide can define their tests, and their students can take these tests on-line. The tests are generated according to teachers' specifications and are adaptive, that is, the questions are selected intelligently to fit the student's level of knowledge. In this way, we obtain more accurate estimations of student's knowledge with significantly shorter tests. By using the computer, larger question databases can be stored, selection algorithms can be performed efficiently, and questions can include multimedia content. The use of Java applets allows the inclusion of executable content in question stem and/or answers, so the student can interact with the system by means of this applet. In this way, new possibilities are added to Computer Adaptive Tests, such as using traditional multiple-choice questions together with questions whose answer is evaluated by the applet itself. The creation of item response theory (IRT) and Rasch models, inexpensive accessibility to high speed desktop computers, and the growth of the Internet, has led to the creation and growth of computerized adaptive
testing or CAT. This form of assessment is applicable for both high stakes tests such as certification or licensure exams, as well as health related quality of life surveys. Gershon, Richard C. (2005) discusses the historical background of CAT including its many advantages over conventional (typically paper and pencil) alternatives and his reflections regarding the future of CAT.

Wise, Steven et.al (2005) investigated several testing methods that may promote examinee effort. One of the advantages of computer-based tests (CBTs) over traditional paper-and-pencil tests is their potential for administering items and tests in innovative ways. The purpose of this study was to investigate the effectiveness of four types of CBTs on the effort and test performance of examinees administered a low-stakes test: a fixed-item test, a computerized adaptive test, a self-adapted test, and a special self-adapted test that was designed to be more "game-like." Seven hundred eleven undergraduate psychology students were randomly assigned to take a low-stakes mathematics test under one of the four CBT types. No differences were found among the test types for test performance, self-reported effort, or response time effort. Although disappointing, the results from this study enhance the understanding of the dynamics underlying examinee effort.

While listing various advantages of E-Portfolio Jain et.al (2006) recommended that assessments became accurate when assisted by Technology than paper – pen reporting and it also increased the comfort of the learner without fear. His appraisal portfolio was structured to assess competence and skill. He defined the purpose was to evaluate student competency as defined by course standards and outcomes.

Georgiadou, Elissavet; Triantafillou, Evangelos; Economides, Anastasios (2007) acknowledged the several advantages of computerized adaptive testing (CAT) over traditional linear test administration, the issue of item exposure control has
received increased attention. Studies have been conducted in order to evaluate the effectiveness of item exposure control strategies for dichotomous scoring, polytomous scoring and testlet-based CAT systems.

Ogan, Amy, et.al. (2009) investigated how to support students in acquiring ill-defined skills of intercultural competence using an online environment that employs clips of feature films from a target culture and found that the addition of pause-predictponder seemed to guide students in acquiring cultural knowledge and significantly increased students' ability to reason from an intercultural perspective. Further analysis of the post test and the post-video viewing discussion found that students in the experimental condition were significantly assisted by the prediction, and were able to maintain a high quality of discussion over time. The research thus establishes that a simple model for intercultural competence activities, enhanced with the novel pause-predict-ponder techniques, is a viable approach to creating tutors in an ill-defined domain, and possibly better suited to the demands of the domain than the standard problem-solving approach embedded in intelligent tutoring systems.

Mullins, Dejana; Rummel, Nikol; Spada, Hans, (2011), studies found positive effects of collaboration on student learning in mathematics, others found none or even negative effects. Collaboration may increase students' learning outcome as it promotes mutual elaboration. If, however, the instructional material is focused on practicing procedures, collaboration may result in task distribution and thus reduce practice opportunities necessary for procedural skill fluency. It was observed that collaboration improved performance during the learning phase in both the conceptual and the procedural condition; however, conceptual and procedural material had a differential effect on the quality of student collaboration: Conceptual material promoted mutual elaboration; procedural material promoted task distribution and ineffective learning
behaviors. Consequently, collaboration positively influenced conceptual knowledge acquisition, while no positive effect on procedural knowledge acquisition was found.

The computer-adaptive, individualized instruction provided by Fehr, Charles; Davison, Mark; Graves, Michael; Sales, Gregory; Seipel, Ben; Sekhran-Sharma, Sarah, (2012), is a vocabulary program addresses a need for efficiency in remediation of vocabulary deficits. Further study is planned to determine whether improved vocabulary performance mediated by this computer assisted language learning (CALL) program might transfer to broader measures of vocabulary knowledge or reading comprehension.

The extent to which learning styles influence the educational process as well as the outcome of elementary-age students in terms of academic achievement was studied by Wilson, Mary, (2012), . The results of this study demonstrate a lack of significant correlation between variables.

2.3.5. Gender Difference in CAT/CAI

In this section various studies related to gender effects in Computer Aided Instruction, testing, and the related researches are discussed:

Gender and Locus of control interactions showed a significant increase in the performance of more externally oriented males and more internally oriented females by the study of Griffin (1990) on the title “Effect of Monitoring Adults during Computer Assisted Instruction in Preparation for the General Education Development Test”. The data were evaluated using analysis of variance and ‘F’ tests.

A study on “The Effect of Time on Computer Assisted Instruction for At-Risk Students” by Salerno (1991) determined if the mathematics achievement of at-risk students using Computer Assisted Instruction (CAI) differed significantly from other groups of students. A one factor ANOVA was conducted on treatments and a two
factor ANOVA was conducted on treatment by gender. Analyses indicated that there was a significant difference between boys having extended CAI time. “A Comparison of Computer Assisted Instruction with Teacher Managed Instructional Practices” by Foley (1994) came out with the major finding that CAI produced increased performance when compared to TMI. Age and sex had no effects on achievement.

“Effects of a Computer Assisted Remediation Program on Basic Skills Mathematics Achievement, Academic Self-Concept, and Locus of Control of Students in a Selected Dropout Retrieval Program in an Urban Setting.” was investigated by Reglin, Gary (1988). The purpose of this study was to determine the effects of a computer assisted remediation program on basic skills mathematics achievement, academic self-concept, and locus of control of students in a dropout retrieval program. A nonrandomized pretest-post test experimental design was used to compare students from urban settings in two groups. Students in the experimental group participated in 12 weeks (60 sessions), 50 minutes per day, of Individualized Manpower Training System (IMTS) instruction plus a 10-minute daily period of computer-assisted instruction (CAI) in mathematics. The only difference between the experimental and control groups was that the experimental group received 10 minutes of CAI in mathematics daily and the control group received 10 minutes of IMTS instruction in mathematics daily. Analysis of covariance indicated no significant difference on any dependent variable for type of instruction and the interaction between sex and type of instruction. The IMTS program with or without CAI had a more significant impact, by t-test, on males' academic self-concept and locus of control.

“The effects of computer-assisted instruction on the achievement test scores of seventh-grade students” by Strohsahl (1994) revealed that students who received instruction in language, which included CAI, performed significantly higher on the
achievement tests in language that students who did not have CAI instruction. Furthermore, male students outscored the female students in language achievement.

Sadatmand (1995) conducted a study on “The effect of Computer-Assisted Instruction on Algebraic Problem solving abilities of Community college students”. Post test mean scores were analysed to determine if a statistically significant difference existed between or within experimental and control groups. A statistically significant difference was found in gender.

DeMars, Christine, (1997), studies gives details regarding the degree of gender differences in mathematics and science appears to vary with the content sub domain. Differences also appear to be greater on items assessing content knowledge than with items measuring reasoning about scientific processes.

Ferguson (1998) worked on the title “The Effect of Locus of Control on achievement and Preference in CAI Tutorial Systems which vary in terms of System control, learner control with coaching and learner control with no coaching” and concluded the following: i) Externals prefer more structure than internals. ii) Externals tend to achieve less than internals iii) Externals follow coaching advice less than internals. iv) Externals make more errors than internals

In their work “On Test and Computer Anxiety: Test Performance under CAT and SAT Conditions”, Shermis, Mark; Mzumara, Howard ; Bublitz, Scott ( 2001) examined differences between computer adaptive testing (CAT) and self-adaptive testing (SAT), including feedback conditions and gender differences in undergraduates. Results of the Test Anxiety Inventory, Computer Anxiety Rating Scale, and a Student Attitude Questionnaire showed measurement efficiency was differentially affected by test condition and also showed significant gender effects.
Clariana, Roy; Wallace, Patricia (2002) described a study that confirmed several key factors in computer-based versus paper-based assessment. Based on earlier research, the factors considered in this study of undergraduates, includes content familiarity; computer familiarity; competitiveness; and gender. Reports results of analysis of variance that showed the computer-based test group outperformed the paper-based test group in their study “Paper-Based versus Computer-Based Assessment: Key Factors Associated with the Test Mode Effect.”

Robinson, Joseph; Lubienski, Sarah; Copur, Yasemin, (2011) have identified gender gaps by Meta analysis. Which, favors males as early as kindergarten or first grade in mathematics, particularly at the top of the achievement distribution. Recent, large-scale studies suggest that teachers actually rate the performance of girls more favorably than the performance of males. Indeed, prior research has revealed that girls tend to exhibit more on-task behavior and positive approaches to learning behavior in schools

Maass, Asja; Kollhorster, Kirsten; Riediger, Annemarie; MacDonald, Vanessa; Lohaus, Arnold (2011), study focuses on the short-term effects of electronic entertainment media on memory and learning processes. The results show that violent content leads to a poorer memory performance for verbal material and to an increased heart rate. The heart rate, however, does not mediate the effect on memory performance. Genders differ regarding their abilities to memorise verbal and visual material, with females showing a better performance (independent of the experimental condition), and for memorizing media-contents, where the males outperform females (also independent of experimental condition). The study supports the assumption that violent and arousing media content has a negative short-term effect on simultaneous information processing and learning and that there are gender-specific media effects.
2.3.6. Students’ Attitudes towards CAT

In this section study related to the Attitude of the learners’ that contribute or affect the performance while administering the test are discussed.

Burts, Diane and others (1991) investigated on “Achievement of Kindergarten Children in Developmentally Appropriate and Developmentally Inappropriate Classrooms.” This study compared end-of-year standardized test scores (CAT) of 204 children in six developmentally appropriate and six developmentally inappropriate kindergarten classes. Particular attention was given to the effects on test scores of the interaction of classroom type with sex, socioeconomic status, and race. Relationships between children's CAT scores and stress scores in the developmentally appropriate class were compared to the same relationships in the developmentally inappropriate class. In the developmentally appropriate class, children who exhibited higher levels of stress during testing scored significantly lower on the CAT average, and on the reading comprehension and language portions of the CAT, than children who exhibited lower levels of stress. These findings suggested that higher levels of stress during standardized testing may negatively affect performance on the test. No significant differences between classroom type and overall test scores were found. Emphasis on academics in developmentally inappropriate classrooms did not result in higher test scores. This finding and previous research that indicated negative consequences of inappropriate curricula suggested that developmentally inappropriate instructional practices were not only potentially damaging to young children's psychological well-being, but that they were also ineffective in promoting achievement in kindergarten children.

In a study on “Individual Differences in Computer Adaptive Testing: Anxiety, Computer Literacy and Satisfaction” Gershon, Richard ; Bergstrom, Betty (1991)
explored the relationship of several individual differences variables to Computer Adaptive Testing (CAT) as compared with traditional written tests. Seven hundred sixty-five examinees took a Computer Adaptive Test and two fixed-length written tests. Each examinee also answered a computer literacy inventory, a satisfaction questionnaire, and a test anxiety survey. Test anxiety was found to be a significant factor in performance on both of the written tests, but not on the CAT test. Anxiety was also found to be a significant factor on several of the items on the satisfaction questionnaire. Overall, significant factors that predicted satisfaction with CAT testing included level of test anxiety, computer literacy, and test length (the CAT test varied in terms of the number of items administered). Results were discussed in terms of the political and practical implications of administering CAT tests as compared to administering traditional written tests. The results also indicated that some of the individual differences variables that had been found to affect performance on written tests were not significant in CAT. Powell, Emily (1992) conducted a research on the title: “Test Anxiety and Test Performance Under Computerized Adaptive Testing Methods.” Little research existed on the psychological impacts of computerized adaptive testing (CAT) and how it might affect test performance. Three CAT procedures were examined, in which items were selected to match students' achievement levels, from the item pool at random, or according to student choice of item difficulty levels. Twenty-four graduate students (5 males and 19 females) at Indiana University (Richmond) were randomly assigned to one of six testing orders formed by the three mastery test approaches and by blocking on native and non-native speakers. While at a computer, students received a description of adaptive testing methods and then a 20-item pre-test anxiety measure. Right after completing each test, students responded to a 10-item in-test anxiety scale, ranked their preferences among
tests, and evaluated their performance on each of the tests. No statistically significant mean differences were found among mean student achievement scores or among in-test anxiety means under the three adaptive testing methods. Students reporting higher anxiety scored significantly higher in the matched-selection test. Those preferring the matched-selection and self-selection tests the most were less anxious during those tests. Instead of actual performance, students' perceptions of how well they did were significantly correlated with preference rankings for the tests. The matched-selection tests required significantly fewer items to reach decisions than did the random-selection tests.

In their paper, “A Comparison of Self-Adapted and Computerized Adaptive Tests.”, Wise, Steven and others (1992) compared the performance of 156 undergraduate and 48 graduate students on a self-adapted test (SFAT)--students choose the difficulty level of their test items--was compared with performance on a computer-adapted test (CAT). Those taking the SFAT obtained higher ability scores and reported lower post test state anxiety than did CAT takers.

In the research work, “The Role of Anxiety in Examinee Preference for Self-Adapted Testing.” by Wise, Steven and others (1993) assessed whether providing examinees with a choice between computerized adaptive testing (CAT) and self-adaptive testing (SAT) affected test performance in comparison with being assigned a CAT or SAT, and evaluated variables influencing examinee choice of either test form. The relative influences of test type and test choice on examinee anxiety were also examined. Students were randomly assigned to SAT, CAT, and choice conditions for an algebra test. Test-related anxiety was assessed with a paper-and-pencil measure in pretests and post tests. It was found that, for students with high mathematics anxiety, providing a choice between CAT and SAT led to significantly higher mean proficiency
estimates, lending support to the hypothesis that examinees could cope with a stressful situation more effectively if they felt that they had some control over the source of the stress. Expected differences in estimated proficiency and post test state anxiety between CAT and SAT conditions were not found, but a strong relationship was seen between examinee test type choice and mathematics anxiety level. Higher anxiety examinees had a greater preference for the control provided by SAT.

“An Investigation of Restricted Self-Adapted Testing” was done by Wise, Steven and others (1993) in which a new testing strategy that provided protection against the problem of having examinees in adaptive testing choose difficulty levels that were not matched to their proficiency levels was introduced and evaluated. The method, termed restricted self-adapted testing (RSAT), still provided examinees with a degree of control over the difficulty levels of their test items. Students were randomly assigned to a computerized adaptive test (CAT), a self-adaptive test (SADT), or RSAT in mathematics. Results indicated no differences between CAT and SADT conditions in terms of mean proficiency and mean post-test state anxiety. The basic RSAT method appeared to hold promise for providing examinees with control over the testing situation, while preventing large mismatches between item difficulty choice and proficiency level. The RSAT procedure should be evaluated empirically.

In a study with the title: “Results of the Administration of the Computerized Grammar, Spelling, and Punctuation Test to the College of Communication Summer 1992 MICOM Group.”, Smith, Nancy and others (1993) reported the results of administering a computerized adaptive version of the Grammar, Spelling, and Punctuation Test (GSP) to students in the College of Communication at the University of Texas, Austin. The computerized adaptive version (CAT) was administered for the first time in June 1992 to 35 prospective students who participated in the College of
Communication's Minorities in Communication (MICOM) program. MICOM was a minority student recruitment program for high-achieving minority students. Statistical comparisons of the two versions conducted in a previous study were used to indicate the relationship between the paper-and-pencil scores and CAT scores. One result of the test administration was the average total number of items administered on the computer version of the test was 60.5 items, as opposed to the 100 items required on the paper-and-pencil version. MICOM students taking the CAT version were asked for feedback about the test and their experiences with microcomputers. Students had favorable attitudes about the CAT test, and most indicated that they would prefer it to the paper-and-pencil version. Students also indicated a great deal of experience with computers.

Computerized-adaptive (CAT) and self-adapted (SAT) music listening tests were compared by Vispoel, Walter; Coffman, Don (1994) in their study “Computerized-Adaptive and Self-Adapted Music-Listening Tests: Psychometric Features and Motivational Benefits” for efficiency, reliability, validity, and motivational benefits with 53 junior high school students. Results demonstrated trade-offs, with greater potential motivational benefits for SAT and greater efficiency for CAT. SAT elicited more favourable responses from examinees.

A Comparison of Fixed-Item, Computerized-Adaptive, and Self-Adapted Testing” by Vispoel, Walter and others (1994) compared Individual Differences and Test Administration Procedures. Vocabulary fixed-item (FIT), computerized-adaptive (CAT), and self-adapted (SAT) tests were compared with 121 college students. CAT was more precise and efficient than SAT, which was more precise and efficient than FIT. SAT also yielded higher ability estimates for individuals with lower verbal self-concepts.
A research work with the title “On Test and Computer Anxiety: Test Performance under CAT and SAT Conditions” was conducted by Shermis, Mark; Mzumara, Howard; Bublitz, Scott (2001). This study of undergraduates examined differences between computer adaptive testing (CAT) and self-adaptive testing (SAT), including feedback conditions and gender differences. Results of the Test Anxiety Inventory, Computer Anxiety Rating Scale, and a Student Attitude Questionnaire showed measurement efficiency was differentially affected by test condition and also showed significant gender effects.

“Testing On Computers: A Follow-up Study Comparing Performance On Computer and On Paper” was an investigation by Russell and Haney (1997). They reported that open-ended test items administered on paper might underestimate the achievement of students accustomed to writing on computers. This study built on Russell and Haney's work by examining the effect of taking open-ended tests on computers and on paper for students with different levels of computer skill. Using items from the Massachusetts Comprehensive Assessment System (MCAS) and the National Assessment of Educational Progress (NAEP), this study focused on language arts, science and math tests administered to eighth grade students. In addition, information on students' prior computer use and keyboarding speed was collected. Unlike the previous study that found large effects for open-ended writing and science items, this study reported mixed results. For the science test, performance on computers had a positive group effect. For the two language arts tests, an overall group effect was not found. However, for students whose keyboarding speed was at least 0.5 or one-half of a standard deviation above the mean, performing the language arts test on computer had a moderate positive effect. Conversely, for students whose keyboarding speed was 0.5 standard deviations below the mean, performing the tests on computer had a substantial
negative effect. For the math test, performing the test on computer had an overall negative effect, but this effect became less pronounced as keyboarding speed increased. Implications were discussed in terms of testing policies and future research.

Christensen (1997), Conducted a study on “Effect of technology integration education on the attitudes of teachers and their students.” The Research Findings are Attitudes toward technology become more positive with “training”, Some attitudes change more quickly than others (anxiety, acceptance, etc.), Teacher attitudes towards technology impact student attitudes, Attitude change is the precursor to new behaviors, Teacher Training is Essential.

An investigation was carried out in introductory astronomy courses comparing pre- and post course administration of assessments using the web and on paper. Overall no difference was seen in performance due to the medium. Compliance rates fluctuated greatly, and factors that seemed to produce higher rates are identified. Notably, email reminders increased compliance by 20% (Bonham, Scott, 2008).

He, Wei; Reckase, Mark, (2008), studies results from both simulation and statistical hypothesis testing are in consistent with schnipke & scrams (1999) in terms that the presence of disclosed items can inflate the final ability estimates in the context of cats by an unacceptably high degree, especially for the low-ability examinees. among all item exposure procedures under study, the a-stratified test with b-blocking, i.e., a design that stratifies the item pool but takes care of the relationship between item difficulty and discrimination parameters, performs the best.

Demirci, Neset (2010) compared undergraduate students' homework performance using a web-based testing system with paper-based, hand-graded one in introductory physics courses and also investigated Students' perceptions about each
method. Overall, it was observed that students' perceptions about the web-based homework system were positive.

Recent research indicated that students' ability to construct evidence-based explanations in classrooms through scientific inquiry is critical to successful science education. Structured argumentation support environments have been built and used in scientific discourse in the literature. The proposed work i.e. “intelligent argumentation assessment system based on machine learning techniques for computer supported cooperative learning” was effective in classifying and improving student's argumentation level and assisting the students in learning the core concepts taught at a natural science course on the elementary school level (Huang, Chenn-Jung, et.al., 2011).

Pea, Celestine, (2012), conducted a larger study, on teachers' beliefs about science teaching, one component looks at how school environmental context factors influence inquiry-based science instruction. Research shows that three broad categories of school environmental factors (human, socio cultural, and design) impact inquiry-based teaching in some way. Results show that few school environmental context factors impact teachers' ability to teach science using inquiry-based methods.

Wilson, Mary, (2012), extended their research to find out, which learning styles influence the educational process as well as the outcome of elementary-age students in terms of academic achievement was studied. The results of this study demonstrate a lack of significant correlation between variables

2.3.7. Learning Time

In this subsection, studies related to response time are discussed:

In a research paper titled as “Adaptive learner Control Provides a more Efficient Frame work for Education (CAI)”, Miller (1996) reported that, when adaptive learner
control was applied students who entered the lesson of study at a level of high performance would finish the lesson in less time than those who entered the course at an average level. Furthermore, those entering at an average level finish the lesson in less time than those who entered the course at a low performance level.

Cilliers, and Others, (1996) determines the extent to which aptitude alone contributes to academic performance with specific reference to performance in the introductory level physics practical. Establishes which specific aspects of aptitude are the most salient predictors of success in the physics practical. Attempt to establish whether familiarity with the language of tuition has an effect on the relationship between aptitude and achievement (MVL).

Van der Linden, Wim. et.al (2003) proposes an item selection algorithm that can be used to neutralize the effect of time limits in computer adaptive testing. The method is based on a statistical model for the response-time distributions of the test takers on the items in the pool that is updated each time a new item has been administered. Predictions from the model are used as constraints in a 0-1 linear programming model for constrained adaptive testing that maximizes the accuracy of the ability estimator. The method is demonstrated empirically using an item pool from an operational, large-scale computer adaptive test.

Gordon Cawelti (2003) stressed that pausing after asking a question resulted in an increase in achievement. Selection of the level of question and thinking time for the question could be added to Test time. It might be considered as the reaction time in the process of Testing necessary to reveal the amount of interaction between the internal and external varieties. CAT score might be considered as the product and the preceding stage process can be known as the Test time.
Chang, Shu-Ren et.al (2005) examined the time different ability level examinees spend taking a CAT on demanding items to these examinees. It was also found that high able examinees spend more time on the pretest items, which are not tailored to the examinees' ability level, than do lower ability examinees. In which Higher able examinees showed persistence with test questions, regardless of the item's difficulty level on operational and pretest questions they answer correctly or incorrectly and found that more able students spend more time on all items, regardless of whether the items are answered correctly or incorrectly.

Several computer-related variables, such as students' access to a computer at home and how often students accessed the Internet were included in the study by Slykhuis, David; Park, John, (2006), Math ability and physics aptitude dominated the models regardless of the method of delivery of instruction. Computer-related variables were included in the model for the online group; however, they were not significant.

Seery, Michael, (2009), conducted a study to find out the role of prior knowledge in first year performance of undergraduate chemistry was examined, considering specifically the context of general student aptitude. Statistical testing shows that there were significant differences between the mean scores of students who have and those who have not prior knowledge of chemistry. Correlation analysis shows a strong correlation between prior knowledge and exam performance, and allows for probing of the role of student’s aptitude. Finally, regression analysis confirms that prior knowledge has a demonstrable influence on future exam performance over and above student aptitude. Two prediction models based on the regression analysis study are proposed.
2.3.8. Learning Rate

Edwards; Norton; Taylor; Weiss; And Dusseldorp, (1975) investigated on the title "How Effective is CAI? A Review of the Research" and reviewed research on the effects of CAI on achievement, retention, and learning rate and its effects on students of different ability levels. CAI as a supplement to traditional, teacher-directed instruction was found to be very beneficial.

Under the title “Computer-Assisted Instruction”, Rapaport, and Savard, (1980) reviewed and synthesized research on the effects of CAI on student achievement, attitudes, and learning rate. Found traditional instruction supplemented by CAI superior to either method alone, and found CAI to be beneficial to student attitudes and learning rates. Ability is predictive of item-level response times for items on the verbal section for both datasets, while item difficulty is predictive of item-level response times for certain sets of quantitative and analytical items. In each case, the regression equations explain more of the variability in the item-level response times when the data were administered under the proportional adjustment scoring rule (Swygert, Kimberly A., 2003).

The study by Giraud, Gerald & Smith, Russel (2005) examines the effect of item response time across 30 items on ability estimates in a high stakes computer adaptive graduate admissions examination. Examinees were categorized according to 4 item response time patterns, and the categories are compared in terms of ability estimates. Significant differences between response time patterns were observed. Highest ability estimates were associated with a response time pattern that was consistent across items, while lowest ability estimates were associated with long response times on items early in the test and short response times late in the test. These results suggest that teaching examinees to manage time effectively can maximize
ability estimates. An alternative interpretation is that more able examinees require less
time to respond to items

Davamani Christober (2006) conducted a study on “Concept Mapping on Mathematics Education through Computer Assisted Instruction.” In this study the operation of internal and external variables was reflected by the rate of interaction to produce the mean achievement score. It was found that the learning rate could influence the process of CAI based learning that resulted in increasing the achievement score through the learning process. Further, mean achievement score, the indicator of learning process in CAI was compared with learning rate in one of the comparisons in the final analysis.

Megala (2007) study on CAT, shows the CAT influences on the learning rate and CAT high achievers have high learning rate and low scorers shows high achievement scores in the post test.

Three cognitive variables--speed of information processing, verbal knowledge, working memory were studied by Rast, Philippe, (2011), and found that all 3 predictors, affected positively initial recall, the asymptotic performance increased with better working memory and faster processing speed, and the learning rate was positively associated with verbal knowledge only. Age did not affect the learning parameters but correlated negatively with working memory and processing speed. The finding of large and reliable individual differences in learning is seen as evidence that the potential for positive change, or plasticity in adulthood is maintained and that it is worthwhile to enhance the determinants of learning or learning itself.

2.4. CONCEPTUAL FRAME OF REFERENCE

From the reviews, it is inferred that both theoretical perspectives and advancements in technology had influenced the direction of research in CAT. The
introduction of the personal computer in the classroom focused on Computer-Assisted Instruction (CAI), which refers to any kind of computer use in educational settings and includes drill and practice, tutorials, simulations, instructional management, supplementary exercises, knowledge assessment, ability estimation, classification of learners, database development, writing using word processors, and other applications. The collection of references is finally sieved for its close relationship with the nature of study undertaken. An attempt is made in classifying them into various categories but, many of them have cross references embracing either areas noted or dimensions marked under different categories in more than one. Collectively these references have helped to precipitate the theme of the study, which is arrived, and the conceptual frame use this as a tool for future self development was made out of the reference work.


Some of the cognitive and affective ways computer technology can be used for assessing students in science include: Computer Adaptive Testing; Panastasiou, Elena c. (2002), Ricardo Conejo, et.al. (2004), Woolnough, Jim, et.al., (2006).
Ricardo Conejo, et.al. (2004) developed a system called SIETTE, and its theoretical bases are Computer Adaptive Testing and Item Response Theory in this, the student can interact with the system by means of this applet. In this way, new possibilities are added to Computer Adaptive Tests, such as using traditional multiple-choice questions together with questions whose answer is evaluated by the applet itself.


The development of the package and the procedures of validation are drawn from many of the studies conducted in developing instructional package, Bergstrom, Betty (1992), Reed&Castle (1997), Lawrence M. Rudner (1998), Mridula D. Ranade, Froelich and Amy Goodwin, Kocakaya, Serhat & Gonen, Selahattin (2010) analyzed and compared the higher, lower-order cognitive skills type physics questions.

The review shows that a wide variety of techniques had been used in designing and developing computer adaptive testing. Since the purpose of present study is different from the earlier studies, as it seeks to develop a Computerized Adaptive Test (CAT), the researcher has designed a new technique, which is not presented before. In computerized adaptive testing (CAT), the computer continuously reevaluates the ability of the student in a test. This fine-tuning is achieved by statistically tailoring the test to the achievement level of each student (Woolnough, Jim, et.al., 2006). The use of learning time calculation is done as recommended by Gabriel (1996), Miller (1996)and Megala (2007) has been given due consideration.

The review has provided the following suggestions to the researcher, for the preparation of the computerized adaptive testing package:

- The test items has to tailor according to each individual student (Ricardo Conejo, 2004)
• Questions used in a Computer Adaptive Test should have one objective of measurement at a time (Reese & Lynda m, 2003)

• The order of presenting questions does not determine the success or failure in answering that question. (Lawrence .Rudner, 1998)

• Correct responses are followed by difficult items.

• Incorrect responses are followed by easier items.

• Item pools with 150-200 items are preferred for satisfactory implementation of CAT. (Weissman, Alexander, 2003)

This review further revealed the variety of tools planned in different studies to achieve the purpose for which the study was conducted. Among these, the researcher has selected the tool for data collection used by Simon (2000) for assessing the Attitude towards Individualization and Technology Preference Tool used by Megala (2007). Pre test mean score as a whole and post test mean score as a whole are being used to indicate the level of gain of learners’ who are not treated as slow learners or high achievers either. Similarly, the achievement of the learner in CAT is to be calculated as the true ability estimate of the learner. Besides learning outcomes, research also examined the effects of CAT on student attitudes, Attitude towards individualization, Aptitude in Physics and preference to other technologies in education. Researchers have also found evidences to suggest that CAT enhances learning rate Rast, Philippe (2011). However, in some research studies, students learned the same amount of material in less time than traditionally instructed Dunkel, Patricia (2000); in others, they learned more material in the same time.

However, most researchers do not specify how much faster students learn through CAT. The review shows that very little work has been done on learning rate. Researcher has adapted an approach used by Megala (2007), which is based on the
harmonic progression to calculate the learning rate of the students; it helps to measure how quickly the concepts are attained by the students’

2.5. SUMMARY OF THE REVIEW

The reference work scanned from 1980 to till date covering the wide range of studies relevant to the theme of research, is practically sufficient.

Observations globally with representations on CAT in science, in terms of Achievement score, CAT score and its effectiveness on intervention. The computerized Adaptive test material was developed with the basis of the field work carried out. The IRT studies and Learning rate are carefully designed covering the most recent areas in educational research such as learning time & Attempt; the two essential parameters still not catching the focus of attention of Individualized Instruction.

Indian references though scanty; but thesis work published like Christobor (2005); Amali Megala (2007), and the researcher’s articles were given appropriate focuses. The summary of the related studies are tabulated and are given in the appendix-7.

2.6. CONCLUSION

The various studies conducted may not be able to give a direct focus, but they act as guidelines to point out the direction for further analysis. The use of diverse tools and techniques in these research studies, made it difficult to derive some common points. However, an attempt is being made to draw some general conclusions, which helped the researcher to give a final shape to the problem identified for the present study. The next chapter deals with Methodology.