3.1 Introduction

Reviewing of the existing literature has been done as one of the processes to progress present research study. It was felt - rather found essential - to search the literature, which addressed following relevant and important aspects:

(a) Studies that have been made in the past in India and abroad that encompassed directly or indirectly the subject of the present research study.

(b) Has the education system in India been geared up to ensure equitability amongst the 'haves' and 'have nots' in India.
(c) What studies have been made in the past that could help draw and establish hypothesis.

(d) The Institutions or bodies that have or may have the role to play in regulating the examination system in India so that their role could be considered for augmentation.

(e) Factors that determine the merit of the people and how these factors can be explored to arrive at inferences to make suitable recommendations as deliverable(s) of the present research.

(f) How can the intelligence be measured and are there any such process(es) in place in India.

(g) Interpretation of test scores and norms; different norms of testing the students.

Discussion in this chapter would indicate that the efforts are made to find responses addressing the above points to help make suitable assumptions and establish hypothesis.

To begin with discussion on studies that have been made in the past in India and abroad that encompassed directly or indirectly the subject of the present research study has been made.
3.2 Status Report On higher Education in India

Prof. Philip. G. Altbach, a prolific writer and well known educationist and an expert of repute in social science while analysing an Indian case in the book titled "Contemporary India, some sociological Perspectives", makes a very lucid, true and apt commentary on Roads to modernisation in India and describes the whole scene on the discussion that follows.

Key ingredient in almost all socio-economic planning is higher education. Without regard to economic system or political organization, the university is considered to be crucial in the process of modernization. Higher education is supposed to do the following amongst many things:

(a) Instil a sense of nationality in countries without basic national unity;
(b) It should provide the advanced technical training necessary for modern industry;
(c) It should educate teachers and agriculturists who can spread modern attitudes and techniques to the countryside;
(d) It should enable the nation to compete in the international marketplace of ideas;
(e) The university should provide national prestige.

There is little data concerning the contribution of the universities to various aspects of public life is available, and even little information concerning the
growth and current status of higher education in India is known. Examination of some of the facts concerning the relationship between higher education and modernization with respect to India against widely held notions about higher education and modernization in India, as in other developing countries, is made to identify most effective role universities are to play.

The organizational structure of the educational institutions that emerged reflected the biases and educational views of the British. The curriculum followed the English patterns and strongly emphasized humanistic studies. Instruction was generally through the English medium and little attention was paid to the Indian context of education. The early Indian colleges had as a goal the creation of “gentlemen” in the English pattern. Barring a few exceptions, standards were below those found in England at that time.

The establishment of universities in India took place in 1858. It systematized higher education and greatly stimulated its expansion. The universities, however, did not substantially change the orientation of the educational system. The curriculum remained basically the same; and the organization of the colleges, traditionally the keystone of the system, underwent few changes. An examination system that provided some unity to the geographically diffused colleges commenced after universities were established. But they did not basically unify the system, since the universities were, at first, only examining bodies and had insignificant teaching and research programs.
The universities were patterned after the University of London. It was felt that the London model, which provided an umbrella institution for many diffused colleges and institutes, would be most suitable for India. The London model was not copied entirely. The Indian universities were not planned to have the same level of instruction. The universities were not specifically intended to contribute to any process of modernization or self-government in India. Implicit in this system was the notion of “trickle down” education in which the benefits of Western style higher education would be given to a very small minority, who would then be counted upon to spread enlightenment to the masses. The universities did provide the ideological basis for nationalism, as well as a pool of trained manpower, which could take over the running of an independent India.

The colleges, despite their impressive quantitative growth, have not kept up in terms of quality. Their curriculum has not been basically modernized, and textbooks and teaching methods remain outmoded. In many areas, the language of instructions has been changed from English to the regional medium. This shift has not been accompanied by a similar change to an “Indian” curriculum. The localization of language, as also of the composition of the student body has meant that some of the cosmopolitanism, which previously was characteristic of the colleges, has been lost. The circulation of talent has broken down at least part of the regionalism in the cities.
"A study of grading systems in different universities for development of normalisation technique to streamline recruitment processes"

"Colleges have opened their doors to a larger number of students from different strata of the population. This has enabled to accommodate a larger number of students from diverse social classes and economic backgrounds. The result of this accommodation has been, almost universally, a lowering of educational standards. University examinations have suffered a decline in standards, as has the general level of college instruction. College teaching, never a well paying occupation in India, has declined in relative remuneration and has also lost some of its prestige. This has meant that qualified graduates will choose an occupation other than college teaching if they have a choice."

The expansion in numbers of seats available in colleges in India, the shift to regional languages as media of instruction, and a general raising of the level of aspirations for the lower middle class and even some workers and peasants has accounted for the demand for expansion of higher education, and has shaped the nature of expansion. Just as higher education has been responsive to the demands of the marketplace in its expansion and development, the universities have also responded to the needs of industry and government and have added new subjects and specialties. The traditional arts and sciences retain their dominant position in terms of size and geographical spread. Government initiative has also helped in the establishment of new institutes and specialties within the universities.
The expansion of higher education in India has not been according to any plan and has not been directly related, for the most part, to the needs of a modernizing economic and social system. Since there has been no plan, it is likely that higher education has not contributed as much as it might have; and certainly its contributions have not had any major direction or goal. The impact of the various official reports on universities, from the Radhakrishnan Report to the recent Education Commission's Report, has been minimal.  

Education at all levels is a state subject, with only minor intervention from the Center; control has been even more difficult. The fact that the educational system has responded to public pressure has meant that the direction of the change has been shaped substantially by the nature of that pressure. Indian families see a college degree as a necessity for social advancement and employment in a society characterized by extreme difficulty in obtaining jobs.

It is true that in the educational sphere the competing interests of many groups such as university officials and teachers, students, public organizations and lobbies, and the government are all crucial in the policy making.

3.3 Role of Higher Education

The development of higher education in post-independence India is related to modernization and social progress. It is clear that the products of the university system i.e. the trained scientists and technicians have played a major role in the impressive growth of Indian industry and science. University trained people man
most of the posts in a bureaucracy. As has been pointed out earlier, there are limitations to the possible roles of higher education in direct social and economic development if the quality and autonomy of the university is to be maintained. Indian university could have taken a more effective part than has been the case in the past.

The Indian university has to play an effective role in solving crucial problems of regionalism and traditionalism in Indian society. It is possible that the universities, with some degree of national consciousness and a reasonably effective network of communications, could provide some sense of national unity in India.

The universities have also to play significant role in eliminating caste, religious, and regional preferences which divide Indian society. The universities could take a strong stand on these questions, could consciously introduce courses into college curriculum, and could conduct research, which would increase understanding and perhaps lead to ameliorative programs. They need to do these things. It appears that the universities are ridden with regional distinctions as other segments of Indian society despite their public commitment to standards of equity and academic freedom.

3.4 Effects of Higher Education²

In order to fully understand the roles which higher education plays in India, it is necessary to discuss its impact on day-to-day life. The universities have provided the trained manpower, which has helped India's technical, industrial,
and administrative advances. A significant proportion emigrates and joins what has become known as the “brain drain.” It is clear that this overproduction of educated manpower is not only detrimental to many of the individuals involved, but is an unnecessary expense for the educational system. The creation of an “un-employed intelligentsia” also has the potential for political and social unrest, particularly in the urban areas, which can have major implications for India’s future. The difficulty of employment for graduates is a major cause for student unrest in India, and has further lowered the prestige of university teaching and has created psychological and other problems for the students involved.

As stated earlier, the university experience has an impact on the students who undergo it. Studies of the results of higher education in terms of political awareness, attitude change, and other variables in Western countries indicate that higher education does have an effect on the attitudes and values of students. While there is less data concerning these matters available in India, it seems clear that there is a substantial impact on students. The fact that an individual’s performance is judged on the basis of results of objective tests, irrespective of considerations of caste, region, etc. is very important.

In India like other developing countries, higher education has been invariably an urban and an urbanizing phenomenon. Expectations concerning jobs and income are influenced by educational factors and by the university subculture as well. Graduates, even in India’s labour surplus economy, expect that their incomes and accompanying social status will be increased as a result of their college or university degree.
Since independence, higher education has become available to wide segments
of the Indian population. Undergraduate education is available to Indians of
almost every class in the cities and to wider groups in rural areas. An
undergraduate degree in India is often of limited use in terms of employment
opportunities, but it does provide an opportunity. With luck and perseverance, a
young person can hope to succeed in raising his social status as a result. The
impact of higher education is positive in terms of modernization/ nation building.

The potential role of higher education is quite substantial. The universities could
contribute effectively in a number of areas. Careful studies of the direction and
needs of high level research and planning could help to channelise manpower
constructively. A limitation on academic growth and emphasis on the maintenance and
improvement of standards in universities could help to uplift the academic system.
Higher education might also play an important integrating role in India. Since the
colleges and universities are the few institutions in India which are similarly organized
throughout the country and which have a common goal, it is at least conceivable that
they could be united into a common task of breaking down the barriers of regionalism,
communalism, and religion.

Unlike other developing countries India has been relatively careful to preserve
academic autonomy. There must be a balance between the needs of the
society and of the economic and social planners on the one hand and of the
academic leadership on the other. If the universities are to contribute to
3.5 Intelligence Distribution: Relation With Status and Strata

Ms. Chitra Siva Kumar stated that broadly speaking, sociologists have used the term education in two senses; namely:

(a) A broad one encompassing the varied processes of socialization that occur in societies; and

(b) A more restricted one referring only to formal instruction given in educational institutions.

In the present research study the term is used in latter sense.

In both developing and developed societies, education is viewed as an important vehicle of social transformation. The problem of social transformation and the role of education in it are crucial. Education is conceived as assisting the society to achieve the goals it has set for itself, that is, goals in the field of technological development, or social goals such as creating a casteless society, or moulding the character of individuals as responsible, socially conscious members of the society.

The Government of India, after Independence, adopted several measures for the educational and social upliftment of the weaker sections of the society. Though these measures have brought about a marginal improvement in their
educational and social position yet they are nowhere near the traditionally privileged sections. In this context, it may be noted that the various measures to reduce inequalities have failed to achieve their objective.

The main objectives of Ms. Chitra’s research under reference were:

(a) To delineate the distribution of the utilization of higher education among the various Hindu castes, and among the various income, educational and occupational strata of Mysore society and to analyse the role of sociological factors in it.

(b) To make clear its implications for the social mobility of individuals and groups.

(c) Compare between some of the results of her study with a few other studies to bring to bear a wider perspective on the crucial dimension of higher education.

The data presented there in respect of caste, income, and educational and occupational background of students was collected in the years 1963-64. Though the information was limited to women students, it may be considered to reflect, in broad terms, the distribution of utilization of higher education among the various Hindu castes and other strata in the total population of the city.

Studies also show how besides caste, other factors such as income, educational and occupational background of individuals, play an important role...
in determining access to higher educational institutions. For instance, Shah found that among the students in the University of Baroda, the largest representation was of the upper income groups.

The theme of social transformation of Indian society has engrossed the attention of sociologists for the past three or four decades, and their writings cover a wide range of themes in that area. Some of them have provided micro-level perspectives of change while others have attempted macro level perspectives.

Myrdal made the following observations on the theme of 'Equality and Democracy' in India and other South Asian countries in his Asian Drama:

"There is a paradox in the South Asian situation; although greater equality has been proclaimed as an immediate practical goal for planning and policy. Marked inequality exists everywhere. The disparity is more striking because, despite more or less successful attempts at planning, economic inequalities have generally not decreased since independence; if anything, they have increased in all the countries of the region. This trend is most apparent in India where planning is taken seriously not only as a theoretical exercise but also as a practical government activity."
Discussing the futility of measures to promote equalisation, Myrdal stated further:

"Under these circumstances equalisation has been attempted through a number of palliative measures. Many supposedly egalitarian measures actually favour middle and upper class groups and discriminate against the masses. We shall find that this is broadly true of various assistance schemes and of land tenancy reform efforts. Government policies in regard to community development, agricultural extension, and co-operation are examples on a large scale. Measures specifically designed to aid the lowest strata in the population have ordinarily been poorly enforced at all."

Ms. Chitra concludes her research work with remarks that in her study she was delighted to note that she did not discover anything to suggest that intelligence is restricted to any particular caste or financially better off students, though their presence in higher classes of education is less in proportion to their actual presence in the population.
3.6 Higher education and UGC

While documenting Indian National Affairs investigations were made to find answers to questions such as how far the educational planning covering the following has served its purpose:

(a) To meet India's manpower requirements; and
(b) To produce enlightened citizens able to constructively relate to the social environment.

Subject of education has remained dynamic since independence. There have been two Education Commissions to lay clear guidelines for educational objectives and goals. There have been other government and social agencies working to realise these goals and this always remained in the deliberative process.

Since the last few decades, the relevance of the educational system in prevalent conditions has been discussed again and again in several forums in order to give a new thrust to the subject. The two-day conference of the vice chancellors in 1982 in New Delhi focussed attention on restructuring of the higher education, the standards of teaching, research and development and the emerging areas of importance.

Inaugurating the 14th seminar on management of educational system under the auspices of the Administrative Staff College of India, the then Union Minister of
State, Mrs. Sheila Kaul, had said “the time had come for developing an appropriate model to manage the universities and make them dynamic instruments of change and development.”

According to Prof. Rais Ahmed, then vice-chairman of the UGC, “the future of the country would be bleak if we failed to change our perception about the education.” This reveals the planners are not oblivious to the changing environment.

Institutions entrusted with the task of imparting higher education and carrying out research projects have to be especially noticed for their output in terms of numbers. The overall situation is that India now is on the world map for making a serious effort on the educational and research front.

Major responsibility for educational targets rests with the Ministry of HRD. Two Commissions have produced very comprehensive documents. The First Commission was appointed in 1949 under Dr. S. Radhakrishnan, an eminent scholar, philosopher and later to be the President of India, as Chairman. The second Education Commission (1964-66) was headed by Dr. D.S. Kothari, then chairman of the University Grants Commission, to advise the Government, on the national pattern of education and on the general principles and policies for the development of education at all stages and in all aspects.
The Second Commission's report submitted to the Government in 1966 is frequently cited in matters of educational significance. The report while being appreciative of the structural work done by the Radhakrishnan Commission in assessing the educational needs of the country examined afresh many educational issues. The Commission's Task Forces were deployed on:

(a) School Education;
(b) Higher Education;
(c) Technical Education;
(d) Agricultural Education;
(e) Adult Education;
(f) Science Education and Research;
(g) Teacher Training and Teacher Status;
(h) Student Welfare;
(i) Research Techniques and Methods;
(j) Manpower;
(k) Education Administration, and
(l) Educational Finances.

The Commission also went around the States and some Union Territories to have discussions with the teachers, educationists, administrators and students in universities, colleges and schools. Added to this were interviews with several distinguished men and women in public life, science, industry and scholars in
"In the rapidly changing contemporary world, universities are undergoing profound changes in their scope, functions and organization and are in the process of rapid evolution. Their tasks are no longer continued in the two traditional functions of teaching and advancement of knowledge. They are assuming new functions and the older ones are increasing in range, depth and complexity."

Broadly, the Commission envisaged the following functions for a modern university:

(a) To seek and cultivate new knowledge, to engage vigorously and fearlessly in the pursuit of truth and to interpret old knowledge and beliefs in the light of new needs and discoveries;

(b) To provide the right kind of leadership in all walks of life, to identify gifted youth and help them develop their potential to the full by cultivating physical fitness, developing the powers of the mind and cultivating right interests, attitudes and moral and intellectual values.

The other important functions mentioned by the Commission are:

(a) To provide society with competent men and women trained in agriculture, arts, medicine, science and technology and various
other professions, who will also be cultivated individuals, imbued with a sense of social purpose;

(b) To strive to promote equality and social justice and to reduce social and cultural differences through diffusion of education; and

(c) To foster in the teachers and students, and through them in society generally, the attitudes and values needed for developing the 'good life' in individuals and society.

In spite of the mammoth efforts of the UGC which has the "statutory responsibility under the UGC Act, 1956 (Act No. 3 of 1956) and the UGC Amendment Act, 1972 (Act No. 33 of 1972) UGC has to take all such steps as it may think fit for the promotion and coordination of university education and for the determination and maintenance of standards of teaching, examinations and research in universities."^10,11

Having discussed some of the social angles related with the present research work it is now attempted to explore some of the central aspects of intelligent behaviour, its measurements etc. The purpose is to address three broad questions about the nature of intelligence in succeeding paragraphs:

(a) What is intelligence, and to what extent is it a unique attribute of the human species?

(b) How can intelligence be measured or evaluated?

(c) What is the nature of the mechanisms that are capable of intelligent behaviour?
3.7 Intelligence Described

Intelligence is easier to recognize than to define or measure. While the word "intelligence" is used in ordinary conversation, and has a dictionary definition, it has no single scientific meaning. No quantitative natural laws relating to intelligence have as yet been discovered. In view of this situation, the concept of intelligence is subject to change as the understanding of human intelligence increases. Further, without a scientific definition, much of the social debate over matters relating to intelligence (e.g. contentions about racial differences with respect to intelligence) cannot be rationally resolved.

A dictionary definition of "intelligence" includes statements such as:

(a) The ability to meet (novel) situations successfully by proper behaviour adjustments; or

(b) The ability to perceive the interrelationships of presented facts in such a way as to guide action towards a desired goal.

The word "learning" is associated with the first statement, and goal-oriented behaviour, problem solving, and understanding with the second.

3.7.1 Attributes of an Intelligent Agent

An intelligent agent is expected to be able to:

- Have mental attitudes (beliefs, desires, and intentions)
• Learn (ability to acquire new knowledge)

• Solve problems, including the ability to break complex problems into simpler parts

• Understand, including the ability to make sense out of ambiguous or contradictory information

• Plan and predict the consequences of contemplated actions, including the ability to compare and evaluate alternatives

• Know the limits of its knowledge and abilities

• Draw distinctions between situations despite similarities

• Be original, synthesize new concepts and ideas, and acquire and employ analogies

• Generalize (find a common underlying pattern in superficially distinct situations)

• Perceive and model the external world

• Understand and use language and related symbolic tools

Some additional attributes of intelligence include reasoning, common sense, planning, perception, creativity, and memory retention and recall.

3.7.2 Attributes related to, but distinct from, intelligence

There are a number of human attributes that are related to the concept of intelligence, but are normally considered distinct from it. These are:

• Awareness (consciousness)
- Aesthetic appreciation (art, music)
- Emotion (anger, sorrow, pain, pleasure, love, hate)
- Sensory acuteness (smell, touch etc.)
- Muscular coordination (motor skills)

3.8 Theories of Intelligence

Theories of intelligence are primarily concerned with identifying the major independent components of intelligent behaviour, and determining the importance of, and interactions between mechanism, process, knowledge, representation, and goals. In particular, such theories address the following issues:

- **Performance theories**: How can one test for the presence or degree of intelligence? What are the essential functional components of a system capable of exhibiting intelligent behaviour?

- **Structural / function theories**: What are the mechanisms by which intelligence is achieved?

- **Contextual theories**: What is the relationship between intelligent behaviour and the environment with which an organism must contend?

- **Existence theories**: What are the necessary and/ or sufficient conditions for intelligent behaviour to be possible?

Quantitative definitions of intelligence range from implicitly defining intelligence as that human attribute which is measured by IQ tests, to assuming that the total information processing capacity of the brain is measured by its size.
(Beyond that needed to support normal body functions). However, the dimension along which definitions of intelligence differ most is the structural (internal) versus the contextual (external). At the structural extreme, intelligence is viewed as the competence of the human (or animal) nervous system to reason, while at the contextual extreme, intelligence is viewed as the ability of an organism to adapt to its physical and social environment. In the latter case, goals, expectations, stored knowledge, and prior experience are as important and relevant as the internal reasoning machinery.

Theories of intelligence are largely dependent on whether one defines intelligence to be a natural phenomenon appearing in living organisms (especially man), or to be an abstract facility with certain specified properties. From a practical standpoint, one might seek what kinds of measurements are needed to predict human performance in specified tasks requiring intelligence.

If intelligence is a highly integrated process, then it is quite possible that a single number, such as an IQ test score, could be a good predictor of a human’s ability to perform in any intellectual task domain. Most psychological theories of intelligence, and intelligence tests that implicitly arise from these theories, assume that intelligence is a composite of a relatively small number of component factors, possibly dominated by a single integrating factor. These theories can be called “performance theories”, since they are based on measurements of performance and make assertions about relationships and
correlations between different tests of performance. Such theories are largely empirical as noted by Butcher.¹⁵

"The study of human intelligence has yielded a large accumulation of knowledge about individual differences, but very little about the basic laws of cognitive functioning. For a concept to be valuable it should have more than purely statistical support, and be more than a blind abstraction from a set of correlated performances."

Structural/ function theories of intelligence are theories that propose certain physical or formal structures as the basis for intelligent behaviour, and then examine the functionality that results. For example, if it is assumed that intelligence is a result of formal logical inference, then one might ask if there are human capabilities that could be shown to be unachievable in the formal system because of limitations inherent in logical reasoning. Logical systems do indeed have limitations, which are not usually ascribed to people.

Finally, there are theories (largely philosophical) about the physical conditions necessary for the mechanization of intelligence. These are called existence theories. For example, there is a school of thought that asserts that intelligence is a non-physical property of living organisms, and cannot be recreated in a machine. Another school believes that intelligence is a functional property of formal systems, and is completely independent of any physical embodiment. This latter viewpoint is the one with which the researcher is concerned.
3.9 Theories of Mind

Generally one discusses the attributes of intelligence and intelligent behaviour describing mechanisms that are capable of achieving such behaviour in both living organisms. Introspectively, there appears to be an “inner entity,” the mind, which views the world through the body’s sensory organs, “thinks,” “understands,” and causes the body to react in an appropriate manner.

A primary concern of philosophy is the attempt to understand the relationship between the internal world of conscious awareness and the external physical world. Plato (428-348 B.C.) held that the mind (psyche) was in charge of the body and directed its movements. Plato spoke of the mind as having both appetitive desires and higher desires, and having also a rational capacity to control, direct, and adjudicate between these two types of desires. Later theories held that man was made of two substances, mind and matter. The theory that the mind and body are distinct, known as “dualism,” was given its classical formulation by Descartes in the seventeenth century; he argued that the universe consists of two different substances: mind, or thinking substance, and matter, which can be explained by science and mathematics. Only in man are mind and matter joined together. His concept was that mind was an immaterial non-extended substance that engages in rational thought, feeling, and willing. Matter conforms to the laws of physics with the exception of the human body, which Descartes believed is causally affected by the mind, and which causally produces certain mental events.
The current dominant school of thought regards mind as being a purely physical phenomenon. Sagan\textsuperscript{15} sums up this view succinctly: \textit{"My fundamental premise about the brain is that its workings what we sometimes call ‘mind’ are a consequence of anatomy and physiology and nothing else."} A similar view by Restak\textsuperscript{16} is based on a belief that signals from the brain will some day be understood:

\begin{quote}
\textit{"Since the development of appropriate technologies, it has become obvious that thoughts, emotions, and even elementary sensations are accompanied by changes in the state of the brain. A thought without a change in brain activity is originated the idea that thoughts obey physical laws and can be characterized as computational processes impossible to understand the “mind,” therefore, it is necessary to understand the brain-how concepts are arrived at, the mechanisms underlying perceptions, memory, the neuro-chemistry of our emotions, and so on."}
\end{quote}

Searle\textsuperscript{17} comments on the mind-body problem: "Mental phenomena, all mental phenomena whether conscious or unconscious, visual or auditory, pains, tickles, itches, thoughts, indeed, all of mental life are caused by processes going on in the brain,"

\textit{Newell and Simon\textsuperscript{18} use the information-processing model.} They view formal logic as a way of capturing ideas by symbols, and the algorithmic alteration of such symbols as leading to mind like activity.
One should not think that all modern researchers look at duality with scorn. World famous neurosurgeon Wilder Penfield\(^9\) doubts that an understanding of the brain will ever lead to an explanation of the mind: “Consciousness of man, the mind, is something not to be reduced to brain mechanisms.” Another example of this point of view is contained in “The Self and the Brain” by Karl Popper and John Eccles\(^20\) an updated plea for dualism, the belief that the brain and the mind are distinct entities.

3.10 Measurement / Evaluation of Intelligence

3.10.1 Assessing Human Intelligence\(^{21,22}\)

As noted above while an intuitive concept of intelligence exists, there is no formal or scientific definition of intelligence that is widely accepted. If intelligence cannot be defined, then it certainly cannot be measured in any precise or comprehensive manner. If intelligence tests do not measure intelligence, what do they measure? The purpose of most of these tests is to predict the future performance of the person being tested with respect to ability to compete or perform in an academic program or in a skilled work task. Whether an “intelligence test” actually does have the required predictive power can only be determined by extensive testing in the specific application area.

There are a number of intelligence tests in widespread use, one of the most popular being the Terman-Merrill revision of the Binet-Simon intelligence scale.
The original Binet-Simon work was performed in the period 1905-1911. Binet insisted on three cardinal principles for using his test:

(a) The scores are a practical device and are not intended as the basis for a theory of intellect. They do not define anything innate or permanent. What they measure is not “intelligence.”

(b) The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking normal children.

(c) Whatever the cause of difficulty in children identified for help, emphasis should be placed on improvement through special training. Low scores should not be used to mark children as innately incapable.

These principles were disregarded, and his scale was used for testing all children. The Binet-Simon test was superseded by Terman’s 1916 standard version, and then by the Terman-Merrill revision of 1937, and by a later revision in 1960. Some of the categories of items found in the 1960 revision are in Table 3.1 below:

<table>
<thead>
<tr>
<th>Obey Simple Commands</th>
<th>Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify object by use</td>
<td>Opposite analogies</td>
</tr>
<tr>
<td>Repeat digits</td>
<td>Pictures alike and different</td>
</tr>
<tr>
<td>Response to Pictures</td>
<td>Memory for sentences</td>
</tr>
<tr>
<td>Repeat digits reversed</td>
<td>Vocabulary</td>
</tr>
</tbody>
</table>
It is interesting to note that the procedure for selecting questions for this test was that the questions had to satisfy certain preconceived notions of what results the test should produce. This is standard practice in all intelligence tests construction. For example, questions that yield systematically higher scores for either boys or girls are eliminated. By use of question selection and scoring procedures, the test was constructed so that for the white American population, biased somewhat towards urban and above-average socio-economic level persons, the scores would have a normal distribution with an average score of 100, and a standard deviation of 16.

Another commonly used intelligence test; the Wechsler intelligence Scale uses separate tests for adults and children. This test is divided in to two main parts, one to test predominantly verbal ability, and second to test performance:

(a) **Verbal Tests.**

- General Information (Who is President of India)
- General comprehension (What would you do if ?... )
- Arithmetic reasoning (Simple mental arithmetic)

<table>
<thead>
<tr>
<th>Memory for stories</th>
<th>Picture completion : man</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find absurdities in pictures</td>
<td>Discriminate animal pictures</td>
</tr>
<tr>
<td>Picture vocabulary</td>
<td>Recognise pictorial objects</td>
</tr>
</tbody>
</table>

Table : 3.1
Remember series of digits forward and backward

Similarities (pairs of words: subject has to tell how they are alike)

Vocabulary (Explain meanings of words)

(c) Performance Tests

- Digit symbol coding (subject must assign digits and symbols to pictures)
- Picture completion (Subject must detect nose missing from the figure)
- Block design. (Construct colour-pattern design in duplication of given patterns)
- Picture arrangement (Subject arranges pictures to tell a story)

Even though the Wechsler and Binet tests have somewhat different categories of questions, they use similar principles of test construction and produce scores that are in reasonable agreement.

Critics have argued that these tests take too narrow a view of intelligence, and that they are based on such dubious assumptions as:

(a) Child is born with fixed or predetermined level of intelligence;

(b) IQ tests can measure this intelligence;

(c) IQ scores will show little variation from early childhood to old age;

(d) The tests employed, relatively unchanged since their introduction in the early 1900s, are good predictors of human performance.
3.10.2 Is Man the Only Intelligent Animal?^{12}

Examination of the attributes of intelligent behaviour presented above and finding examples of superior animal performance in each of the attribute categories reveals that man is not the only intelligent animal. Until recently it was believed that only man, of all animals, could produce (as opposed to understand) structured linguistic phrases to communicate meaning. Experiments have demonstrated that chimpanzees can learn American Sign Language (ASL) and can learn to assign word meanings to physical tokens.

In a related sense, work by Gordon Gallup^{23} addresses the question: “Do minds exist in species other than men?” Gallup defines “mind,” “consciousness,” and “self-awareness” to mean essentially the same thing. His operational test for self-awareness is that an organism can identify itself in a mirror; for example, a child can recognize his reflection at approximately a year and half to two years of age. Gallup discovered that while humans, chimpanzees, and Orang-utans can learn to recognize themselves in mirrors, no other primates could!

3.11 The Mechanism of Intelligence^{12}

In most normal people, the left hemi-sphere of the brain is specialized to deal with tasks amenable to a sequential paradigm. These include language understanding and production, logical reasoning, planning, and time sense. The right hemisphere of the brain is more competent to deal with spatial tasks and tasks requiring a global (gestalt) synthesis. These include comparing and
identifying visual imagery. (There is evidence to support the surprising discovery that mental images are neither generated nor manipulated by the normal sensory based visual system; a module in the left hemisphere, but not language based, appears to provide the necessary competence. There is no similar module in the right hemisphere\textsuperscript{24}) visual and analogical reasoning (including, perhaps, dreaming) and body sense and coordination.

Some of the evidence supporting the concept of specialization of the two brain hemispheres with respect to the gestalt and sequential paradigms has come from split-brain experiments with subjects who have had brain surgery to control epilepsy. The connection between the right and left hemispheres is severed so that signals no longer flow between the hemi-spheres. By examining the subjects of such experiments, it has been found that the human brain can support two separate and distinct “personalities,” one in each hemisphere. The philosophical implications of this finding are rather staggering and are still being investigated.

It would now be in order to turn to another aspect in the context of the present review of literature and that is how tests have to be conducted and with what norms these be evaluated.
3.12 Interpreting Test Scores and norms

Test performance is most meaningful when some basis for comparison is available. Norms provide clearly defined reference groups for this purpose. Interpreting test scores with the aid of norms requires:

(a) An understanding of the various methods of expressing the scores.

(b) The ability to judge what the norm group actually represents. In addition, proper interpretation calls for a frank awareness of the limitations of all test scores.

Test interpretation would be greatly simplified if test scores could be expressed on scales like those used in physical measurement. It is well known, for example, that 1.5 Meter means the same height whether one is talking about the height of a boy or a picket fence. A 120 Kgs. player weighs exactly twice as much as a 60 Kgs. singer. This ability to compare measurements from one situation to another, and to speak in terms of “twice as much as” or “one third as long as,” is made possible by the fact that these physical measures are based on scales which have a true zero point and equal units. Ten Kgs. indicates the same weight to the doctor, the grocer, the farmer, and the housewife.

The properties of physical measuring scales, with which all are so familiar, are generally lacking in educational measurement. A student who receives a score of zero on a
history test does not have zero knowledge of history. There are probably a large number of simple questions he could answer which were not included in the test. A true zero point in achievement, where there is no achievement at all, cannot be clearly established. Even if it could, it would be impractical to start from that point each time the test is conducted. What one does in actual practice is to assume a certain amount of basic knowledge and measures from there. This arbitrary starting point, however, prevents one from saying that a test score of zero indicates "no achievement at all". Since one is never certain how far the zero score on one's test is from the true zero point (i.e., the point of "no achievement at all") test scores must be interpreted in relative rather than absolute terms.

The interpretation of test results is additionally handicapped by the inequality of one's units of measurement. Sixty items correct on a simple vocabulary test does not have the same meaning as sixty items correct on a more difficult one; nor do either of the scores represent the same level of achievement as sixty items correct on a test of arithmetic, science, or study skills. These test items simply do not represent equal units like those of meters, kilograms and minutes.

To overcome the lack of a definite frame of reference in educational measurement, various methods of expressing test scores have been devised. As will be seen shortly, the methods vary considerably in the extent to which they provide units, which have uniform meaning from one measurement to
another. Much of the difficulty in the interpretation and use of test results arises from the fact that there are so many different scoring systems each with its own peculiar characteristics and limitations.

Having dwell on measurements of intelligence it is worthwhile now to discuss how test scores and norms of tests can be set and measurements done.

### 3.13 Types of Test Scores and Norms

#### 3.13.1 Raw Scores

If a student responds correctly to 65 items on an objective test in which each correct item counts one point, his raw score will be 65. Although a raw score provides a numerical summary of students' test performance, it is uninterpretable without further information. It cannot be determined whether it is high or low unless one has some basis to compare. When using informal classroom tests generally the total number of items in the test is compared with the scores obtained by the student's classmates.

Comparing a student's test score against the total number of items in the test makes it possible to determine the value of the score in terms of the percentage of items answered correctly. Score of 65 in arithmetic, for example, would be high if there were only 70 items in the test (93 per cent correct) but low if there were 130 items (50 per cent correct). In general achievement tests, where there is no absolute standard of mastery, "percentage correct" is rather meaningless.
The aim in constructing these tests is to obtain items near the 50 per cent level of difficulty in order to obtain maximum discrimination between high and low achievers.

Scores on informal classroom tests are most meaningful when compared with the scores of other students in the same classroom group. By placing the test scores in rank order, for example, one can determine whether a particular score is third from the top, about average, or one of the lowest scores in class. By interpreting a student's test score in terms of his position in the group, one circumvents the problems arising from errors in judging the difficulty of test items. The fact that a test is relatively easy or difficult for the students does not alter interpretation of test scores in terms of relative performance. If a student's test score is third from the top in a classroom group of thirty students, it is a high score whether it represents 90 per cent of the items correct or 60 per cent correct.

3.13.2 Derived Scores

While raw scores can be used directly for some classroom purposes, their use is restricted in two important ways. Firstly, a raw score is meaningless by itself; it is difficult to interpret and use beyond the immediate situation. To make a raw score of 80 interpretable, for example, information concerning the nature of the group tested and the distribution of scores obtained must accompany it. Secondly, raw scores on different tests cannot be compared directly. If a student obtains a raw score of 55 in spelling and 35 in arithmetic, for example,
there is no basis for determining in which subject he has better achievement. To compare achievement on different tests, unit of measurement is needed.

Derived scores provide units, which approximate the uniformity desired in test scores. A derived score is a numerical report of test performance in terms of the student's relative position in a clearly defined reference group. Converting raw scores to derived scores is simple with standardized tests.

The most common types of derived scores are grade equivalents, age equivalents, percentile ranks, and standard scores. The first two describe test performance in terms of the group in which a student's raw score is just average. The last two indicate the student's relative standing in a particular group of which he is a member, or desires to become a member. The specific meaning of each type of score is given in Table 3.2.

3.14 Norms

Tables of norms included in test manuals merely present scores earned by students in clearly defined reference groups. The raw scores and derived scores are presented in parallel columns so that the conversion to derived scores is easily made. These scores do not represent especially good or desirable performance, but rather "normal" or "typical" performance. They were obtained at the time the test was standardized by administering the test to representative groups of students for whom the test was constructed. Thus,
they indicate the typical performance of students in these standardization groups and nothing more. They should not be viewed as standards, or goals, to be achieved by other students.

Test norms answer questions such as the following:

(a) How does a student's test performance compare with that of others.

(b) How does a student's performance on one test (or subtest) compare with his performance on another test?

(c) How does a student's performance on one form of the test compare with his performance on another form of the test, administered at an earlier date?

These comparisons of test scores make it possible to predict a student's probable success in various areas, to diagnose his strengths and weaknesses, to measure his educational growth, and to use the test results for various other instructional and guidance purposes. Such functions of test scores would be severely curtailed without the use of the derived scores provided by test norms.

Summary of the most common types of test norms is presented in Table 3.2.

To interpret and use test results effectively, one needs a good grasp of the characteristics, advantages, and limitations of each of these norms. Therefore each is described in considerable detail in the succeeding paragraphs.
Table 3.2: Most common types of test norms

<table>
<thead>
<tr>
<th>Type of Norm</th>
<th>Name of Derived Score</th>
<th>Meaning in terms of Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade norms</td>
<td>Grade equivalents</td>
<td>Grade group in which student's raw score is average</td>
</tr>
<tr>
<td>Age norms</td>
<td>Age equivalents</td>
<td>Age group in which student's raw score is average</td>
</tr>
<tr>
<td>Percentile norms</td>
<td>Percentile ranks</td>
<td>Percentage of students in the reference group who fall below student's raw score</td>
</tr>
<tr>
<td>(Percentile scores)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Score Norms</td>
<td>Standard Scores</td>
<td>Distance of student's raw score above or below the mean of the reference group in terms of standard deviation units.</td>
</tr>
</tbody>
</table>

3.14.1 Grade Norms

Grade norms are widely used with standardized achievement tests, especially at the elementary school level. They are based on the average scores earned in
each of a series of grades and are interpreted in terms of grade equivalents. For example, if students in the standardization group who are beginning the fifth grade earn an average raw score of 24, this score is assigned a grade equivalent of 5.0. Tables of grade norms are made up of such pairs of raw scores and their corresponding grade equivalents.

Grade equivalents are expressed in two numbers; the first indicates the Year and the second the month. Grade equivalents for the fifth grade, for example, range from 5.0 to 5.9. This division of the calendar year into tenths assumes little or no change in test performance during the summer vacation months.

It should be especially noted that grade norms indicate the average performance of students at various grade levels. For any particular grade equivalent, 50 per cent of the students in the standardization group are above this norm and 50 per cent are below. Consequently, one should not interpret a particular grade as something all students should attain. If half of the students are above norm and half are below, one may conclude that the students compare favourably with the students in the norm group. Whether this is good or bad depends on a number of factors, such as the ability of students, the extent to which the learning outcomes measured by the test reflect curriculum emphasis, and the quality of the educational facilities. If students were with above average ability under conditions comparable to those of schools in the norm group, then merely matching the norm would be the cause for concern. On the other hand,
if students or educational facilities are inferior to those in the norm group, reaching the norm might call forth considerable pride. In any case, it is well to remember that the norm is merely an average score made by students in the standardization group. As such, it represents the typical performance of average students in average schools and should not be considered a standard of excellence to be achieved by others.

The popularity of grade norms is largely due to the fact that test performance is expressed in units that are apparently easy to understand and interpret. To illustrate, assume that the following grade equivalents obtain for Suresh, who is in the middle of the fifth grade.

(a) Arithmetic 5.5
(b) Language 6.5
(c) Reading 9.0

In examining these scores it is evident that Suresh is exactly average in arithmetic, one year advanced in language, and three and a half years advanced in reading. Grade equivalents provide a common unit with which all are familiar. The difficulty is that this familiarity leads to interpretations, which are misleading those who are unaware of the numerous limitations of grade norms. In fact, the limitations are so severe that over the years test specialists have made a concerted effort to have them replaced by more suitable scores.
3.14.2 Limitations of Grade Norms

The most serious limitation of grade norms is that the units are not equal on different parts of the scale, or from one test to another. A year of growth in arithmetic achievement from grade 4.0 to 5.0, for example, might represent a much greater improvement than an increase from grade 2.0 to 3.0 or grade 8.0 to 9.0. Thus, being advanced or retarded in terms of grade units has a different meaning on different parts of the grade scale. A student who earns a grade equivalent to several grades above his grade placement may be demonstrating vastly superior achievement or performance just slightly above average.

One reason that grade norms provide unequal units is that growth in school subjects is uneven. At grade levels where educational growth is rapid, grade units indicate large differences in achievement, and where growth slows down grade units correspond to very small differences. This difficulty is further complicated by the fact that patterns of growth vary from subject to subject so that these grade units stretch and contract at different points of the scale for different subjects. In general, grade equivalents provide units, which are most comparable when used at the elementary level in those areas receiving relatively consistent instructional emphasis that is, arithmetic, language skills, and reading.

A further limitation of grade norms is that high and low grade equivalents have dubious meaning. Raw scores corresponding to grade equivalents below grade
3 and above grade 9 are usually estimated (by extrapolation) rather than determined by measurement. This does not correspond to the achievement of students in any particular group. Estimating these grade equivalents is frequently necessary because the younger students do not have the needed skills to take the test and because growth in the basic skills tends to level off in the eighth and ninth grades. In interpreting grade equivalents at the extremes, therefore, it is well to keep in mind that they do not represent the actual performance of students at these levels.

The lack of equal units and the questionable value of extreme scores are especially troublesome when comparing a student's relative performance in different subjects. First, note that the reading scores range from a grade equivalent of 3.0 to 10.0, and that the language and arithmetic scores cover a much more restricted range of grade equivalents. This is a typical finding with achievement batteries, which measure the basic skills. These mark the grade equivalent earned by a student on each of the tests. Note that he is at or near the top of each distribution.

On inspecting student's grade equivalents only, it is hard to resist the conclusion that he is far more superior in reading than he is in arithmetic. After all, he is only one year above his grade level in arithmetic, but four years advanced in reading. These differences are quite impressive. They are also misleading since they are largely due to the inequality of grade units from one
test to another. When student's performance on the tests is compared in terms of the percentage of the group falling below him (percentile rank), it is apparent that his performance is identical on these two tests. In other words, when compared to a group of sixth-grade students he holds the same relative position in arithmetic and reading. The range of grade equivalent scores characteristically varies from one type of test to another, resulting in unequal units and distorted comparisons between tests.

Another common misinterpretation of grade norms, although not due to weaknesses in the scoring system itself is to assume that if a student earns a certain grade equivalent score in a subject, he is ready to do work at that level. The grade equivalent score of 6.0 may represent nothing more than a thorough mastery of language skills taught in the first four grades. Thus, grade equivalents should never be interpreted literally. At best, they are only rough guides to level of test performance. Students at different grade levels who earn the same grade equivalent score are apt to be ready for quite different types of instruction.

In summary, grade norms are based on the average performance of students at various grade levels. They are widely used at the elementary school level, largely because of the apparent ease with which they can be interpreted. Grade equivalent scores are based on units, which are typically unequal. This can lead to misinterpretations and tends to limit the usefulness of the test results.
general, grade norms are most useful for reporting growth in the basic skills during the elementary school period. They are least useful for comparing a student’s performance on different tests. For whatever purpose grade norms are used, inequality of grade units must be considered during interpretation of the results.

3.15 Qualities Most Desired in Norms:

3.15.1 Test norms should be relevant. Test norms are based on various types of groups. Some represent a national sample of all students at certain grade or age levels while others are limited to a given region or state. The variety of types of groups available for comparison makes it necessary to study the nature of the norm sample before using any table of norms. Only to compare one’s students with a general reference group in order to diagnose strengths and weaknesses in different areas, national norms are satisfactory. One should ask for these if these norms are appropriate for the students being tested and for the decisions to be made with the results?

Here the main concern is with the extent to which one’s students are similar to those in the norm population on such characteristics as scholastic aptitude, educational experience, and cultural background. The more closely one’s students approximate those in the norm group, the greater is the certainty that the notional norms provide a meaningful basis for comparison.
3.15.2 **Test norms should be representative.** Once it is ascertained that a set of test norms is based on a group with which comparisons are desired, it is appropriate to ask whether the norms are truly representative of that group. Ideally, the norms have to be based on a random sample of the population they represent. As a minimum, it necessitates that all significant subgroups of the population be adequately represented (Stratified sampling).

For national norms, it is desirable to have a proper proportion of students from such subgroups as the following: boys and girls, geographic regions, rural-urban areas, socio-economic levels, racial groups, and schools of varying sizes.

In evaluating test norms, it is easy to be misled by the size of the norm sample. Size of the norm sample provides an insufficient criterion for judging the adequacy of norms.

Whether appraising national or special-group norms, one should always favour the carefully chosen, representative sample over the larger sample based on availability of results. This requires going beyond the titles on the norm tables and making a careful study of the procedures used in getting the norm sample.

3.15.3 **Test norms should be up to date.** One factor that is commonly neglected in judging the adequacy of norms is whether they are currently applicable. With the rapid changes that are taking place in education, one can expect test norms to become out of date much sooner than they did when the
curriculum was more stable and there was less emphasis on accelerated programs. The remarks by Cronbach: 27

"Test norms become obsolete and need to be checked periodically. Research on the Wechsler intelligence tests, for example, suggests that the scores of adults are, on the average, higher than those of similar age groups a decade ago. These changes may be attributed to an increasing level of education."

3.15.4 **Test norms should be comparable.** It is necessary, or desirable, to compare directly scores from different tests. This is the case when profile comparisons of test results are to be made to diagnose students' strengths and weaknesses, or compare aptitude and achievement test scores to detect underachievers. Such comparisons can be made precisely only if the norms for the different tests are comparable. Best assurance of comparability is obtained when all tests have been normed on the same population. This is routinely done with the tests in an achievement battery, and some test publishers also administer a scholastic aptitude test to the same norm group. Whenever the scores from different tests are to be compared directly, the test manuals should be checked to determine whether the norms are based on the same group, if not, they have to be made comparable by other means.

3.15.5 **Test norms should be adequately described.** It is difficult to determine if test norms provide a meaningful basis of comparison unless
something about the norm group and the norming procedure used is known.

The type of information required to find in a test manual includes the following:

(a) Method of sampling.

(b) Number and distribution of cases included in the norm sample.

(c) Characteristics of the norm group with regard to such factors as age, sex, race, scholastic aptitude, educational level, socio-economic status, types of schools represented, and geographic location.

(d) Extent to which standard conditions of administration and motivation were maintained during the testing.

Date of testing including whether it was done in the rainy or spring season.

Other things being equal, one should always favour the test for which one has detailed description and other relevant factors known. Such information is needed to judge the appropriateness of test norms for any particular purpose.

3.16 Cautions In Interpreting Test Score

Interpreting test scores with the aid of norms requires an understanding of the type of derived score used and willingness to study carefully the characteristics of the norm group. In addition, however, one needs to keep in mind the following general cautions, which apply to the interpretation of any test score:

(a) A test score should be interpreted in terms of the specific test from which it was derived. Neither two scholastic aptitude tests nor achievement tests...
measure exactly the same thing. Achievement tests are especially prone to wide variation. For example, one arithmetic test might be limited to simple computational skills while another contains a large number of reasoning problems. Similarly, one science test may be confined largely to items measuring knowledge of terminology while another with the same title stresses the application of scientific principles. With such variation it is misleading to interpret a student's test score as representing general achievement in any particular area.

(b) A test score should be interpreted in the light of all relevant characteristics of the student. Test performance is influenced by the student's aptitudes, educational experiences, cultural background, emotional adjustment, health, and the like. Consequently, when a student performs poorly on a test, it is desirable to first consider the possibility of cultural deprivation, a language handicap, improper motivation, or similar factors, which might have interfered with the student's response to the test. If the test is an achievement test, one must, of course, also take into account the student's scholastic aptitude. A low ability student performing two years below his grade level might be progressing at a rate satisfactory for him. On the other hand, a bright student performing two years beyond his grade level might be achieving far short of his potential.

(c) A test score should be interpreted in terms of the type of decision to be made. The meaningfulness of a test score is determined to a
considerable extent by the purpose of it. For example, an IQ score of 100 would have different meanings if one were selecting students for a mentally retarded class, predicting achievement in high school, or trying to decide whether a student should be encouraged to go to college.

(d) A test score should be interpreted as a band of scores rather than a specific value. Every test score is subject to error and it must be allowed for interpretation. One of the best means of doing this is to consider a student's test performance as a band of scores one standard error above and one below his obtained score. For example, if a student earns a score of 56 and the standard error is 3, his test performance should be interpreted as a band ranging from score 53 to score 59. One has to make allowances for these error bands surrounding each score. This will prevent one from making interpretations, which are more precise than the test results warrant.

(e) A test score should be verified by supplementary evidence. When interpreting test scores, it is impossible to determine fully the extent to which the basic assumptions of testing have been met or to which the conditions of testing have been precisely controlled. Only protection against such errors is to place little reliance on a single test score. As Cronbach\(^{27}\) has noted:

"The most helpful single principle in all testing is that test scores are merely data on which to base further study. They must be
coordinated with background facts, and they must be verified by constant comparison with other available data."

After having discussed the test norms and their correct interpretation and how these could be implemented one would do well now turn to more focused subject of testing of academic performance and scores of the students.

3.17 Academic Testing

The measurement of learning outcomes by means of tests of knowledge and understanding is fundamental in education for both shaping the teaching-learning process and certifying student mastery of the materials to be learned. In practice, one of the principal goals as educators is the development of tests that measure the facets of learning with the greatest precision and verity. The information value and, therefore, the usefulness of any academic measurement are inversely proportional to the amount of error it contains.

During the process of searching the mechanism of testing the academic standards of the students in any education institution it was revealed that there are two approaches namely:

(a) Classical Test Theory Approach

(b) Item Response Theory Approach
3.17.1 Classical Test Theory (CTT) Approach

At the college level, testing for knowledge and understanding in the context of a specific course involves administering the same set of test items or questions to all the students enrolled in the course, usually at the same sitting for the examination. Tests administered in this fashion are referred to as fixed-length tests. A very popular model for tests of fixed length is the classical test theory model.

CTT is best suited for traditional testing situations, either in group or individual settings, in which all the members of a target population, are, administered the same or parallel sets of test items. These item sets can be presented to the examinee in either a paper-and-pencil or a computer format. Regardless of format, it is important for the measurement of individual ability that the items in each item set have "difficulties" that match the range of ability or proficiency in the population. In addition, precise estimation of individual ability requires the administration of a "large enough" number of items whose difficulty levels narrowly match the individual's level of ability or proficiency. For heterogeneous populations these requirements of the "fixed length" test result in inefficient and wasteful testing situations that are certainly frustrating to the examinees.

CTT is a relatively simple model for mental testing which is applicable to and widely used for the construction and evaluation of fixed-length tests. Here only the principles and results of CTT needed to provide the teacher with an understanding of measurement error and its consequences are covered.
The result after conducting CTT may not be conclusive if the questions are such that the student does not know the answer(s) to only these questions and may otherwise be quite intelligent. Such a drawback makes CTT not so useful.

Since 1950's models for mental tests began to appear that addressed these problems of CTT and exploited the emergence of computing technology. Powerful feature of newer testing models was the ability to choose test items appropriate to the examinee's level of proficiency during the testing session. Tailoring the test for the individual in real time was used. More popular and well developed of these models make up a family of mathematical characterizations of an examinee's test responses known as Item Response Theory (IRT). Although difficult to implement in practice, IRT is the formulation of choice for modern testing.

### 3.17.2 Item Response Theory (IRT)³⁰

IRT also begins with the proposition that an individual's response to a specific test item or question is determined by an unobserved mental attribute of the individual. Each of these underlying attributes, most often referred to as latent traits or abilities, is assumed to vary continuously along a single dimension usually denoted #. Under IRT, both the test items and the individuals responding to them are arrayed on # from lowest to highest. The position of person i on #, denoted #i, is usually referred to as the person's ability or
proficiency. The position of item \( j \) on \( \theta \), usually denoted \( b_j \), is termed the item's difficulty. Intuitively, it is expected that the probability of a correct response to the \( j \)-th item to increase monotonically as \( \theta - b_j \) increases.

In terms of binary scored test items, i.e., items on which responses are designated either correct or incorrect, all IRT models express the probability of a correct response to a test item as a function of \( \theta \) given one or more parameters of the item. The three most commonly used IRT models are the one-parameter, two-parameter and three-parameter logistic models.

**Estimation of Item and Ability Parameters**

If the item parameters of the IRT model are known, then the estimation of ability for a sample of examinees is rather straightforward using the method of maximum likelihood. It is common practice to use estimates of the item parameters derived from previous item calibration studies when estimating \( \theta \).

**Item and Test Information**

How precise is the measurement of an individual's ability? This is the fundamental question that any theory of mental testing must address. A common statistical index of error in the estimation of a parameter is the variance of the estimator. In the present context, the larger the variance the larger the error associated with it. Conversely, the smaller the variance the
greater the precision in estimating ability. It is customary when using an IRT model to express the variance as the reciprocal of the information function.

Turning first to the item information function, it is found that this function tends to be symmetric around a maximum score attained. This makes sense in that test items are either too easy or too difficult for an examinee with ability to be less informative than items around . For both the two- and three-parameter models the maximum value of item information is a function of item discrimination. For all models, including the three-parameter model, items having large values are more informative than items with smaller values.

Turning next to the test information function, first note that the items contribute independently and, therefore along with the fact that is defined for each point on the ability scale, item independence means that items can be chosen as adaptive testing to optimise the estimation.

In contrast to Classical Test Theory test information or the standard error of measurement for a fixed set of items varies as a function of ability.

3.17.3 Test Construction

The traditional approach to test development based upon CTT centers upon two item statistics: item difficulty ($p = \text{the proportion of examinees choosing the correct response}$) and item discrimination ($r = \text{the correlation between passing}$

Chapter 3 80
the item and some measure of ability such as the total test score). In general, items are selected whose p values generate the desired test score distribution and that have high item total test score correlations. The final set of items selected should produce a test score, i.e., ability estimate having high reliability. Test score reliability is used to estimate the error in individual test scores.

CTT approach to test development has several weaknesses. For one, the estimates of item difficulty, item discrimination, and reliability are sample specific. For another, error in examinee test scores is assumed to be constant across all values of ability. In contrast to CTT, estimates of IRT item parameters are sample invariant and IRT estimates of error vary as a function of ability. The price of these features of IRT model is requirement of very large sample sizes.

The IRT approach to test development centers on the item and test information functions. This seems reasonable that items should be selected based upon the amount of information each contributes to the amount of information of the test as a whole. Once the items have been calibrated, i.e., item parameters have been estimated, item selection is rather straightforward.

(a) Based upon the purpose of the test, e.g., selecting students for a specific academic treatment, specify a target test information function. This is the test information function for an optimal set of test items.

(b) Next, select test items with item information functions that meet the requirements of the target function.
(c) Compute the test information function as each item is added to the test.
(d) Continue adding items to the test until the computed test information function approximates, in some acceptable sense, the target test information function.

3.18 Some Techniques of Normalization In Vogue

Efforts were made to search if any normalization techniques are already in vogue both in India and abroad. The techniques so located are being introduced in the succeeding few paragraphs.

3.18.1 Law School's (USA) Grade Normalization Policy

In order to achieve fairness and consistency, grades in all courses are "normalized," which means that the average and the distribution of grades are controlled, following the Law School's historic grading patterns.

First-Year Normalization

Grades in all first-year courses are entered into a computerized normalization program (The algorithm used for normalization is not known since the same is not disclosed). Some minimal deviation from the computer-generated grades is allowed. In the rare instances when second year students are enrolled in first-year courses, the grades for the first year students are normalized without including the second year students.
Upper-Division Normalization

Instructors in upper division courses have a choice between submitting their grades to be normalized by a program designed for those courses, and normalizing the grades themselves. The median grade for upper-division courses is the average of the cumulative grade point averages for the students enrolled in the class (and taking the class for a numerical grade). Upper-Division course grades must meet the following criteria:

The target median will be set by the average of the cumulative grade point averages of all numerically graded students. An instructor may increase or decrease the target median by one point in courses with 11-20 numerically graded students, or by two points in courses with 1-10 numerically graded students. Instructors may not increase or decrease the target median in courses with 21 or more numerically graded students.

For all upper division courses, at least 50% of the grades must be within a range of plus or minus five points from the median for the course, and at least 85% of the grades must be within a range of plus or minus ten points from the median for the course. Table 3.3 below display the distribution of grades produced by the computer program for a course with 100 students and 78 or 3.2 as median. The percentages indicate the likelihood of receiving particular grades.
As the table demonstrates, very high grades (87 or 4.1 and above) are given sparingly. The Academic Affairs Committee strongly recommends that such grades be given only to reflect truly extraordinary performance. The Associate Dean and the Registrar review the distribution of grades in all courses normalized by the instructors to ensure that the grades comply with the normalization rules.

**Table 3.3**: Normal Distribution of Grades for a Class of 100 Students with a Median of 78

<table>
<thead>
<tr>
<th>Grades Assigned</th>
<th>Probability of Receiving that Grade or Lower</th>
<th>Probability of Receiving Exactly that Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>66 or 2.0</td>
<td>0.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>67 or 2.1</td>
<td>0.4%</td>
<td>0.2%</td>
</tr>
<tr>
<td>68 or 2.2</td>
<td>0.7%</td>
<td>0.4%</td>
</tr>
<tr>
<td>69 or 2.3</td>
<td>1.5%</td>
<td>0.7%</td>
</tr>
<tr>
<td>70 or 2.4</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>71 or 2.5</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>72 or 2.6</td>
<td>8%</td>
<td>3%</td>
</tr>
<tr>
<td>73 or 2.7</td>
<td>12%</td>
<td>5%</td>
</tr>
<tr>
<td>74 or 2.8</td>
<td>18%</td>
<td>6%</td>
</tr>
<tr>
<td>75 or 2.9</td>
<td>26%</td>
<td>8%</td>
</tr>
</tbody>
</table>
### Participation Points

Faculty may assign points to students to reflect commendable class participation. Faculty who have chosen to do so should have announced their policy on participation points in the registration materials or during the first week of class and should have explained whether points are given in every case or only when the examination falls below the oral (in class) performance. Faculty should also have explained the maximum number of points to be awarded. In
most instances, instructors awarding participation points have assigned from one to three points; it is very rare for more than three points (on the 65-90 scale) or 3 points (on the 4-point scale) to be given for participation.

Relief Points

In all courses in which the grades are normalized by computer, the Registrar provides the instructor with a list showing the distribution of the grades for review before the grades are considered final. At that time, the instructor may make "relief point" adjustments in the grades to bring them into line with the performance of students in the course. These "relief points" are especially intended to permit adjusting the tail ends of the distribution when the computer program has raised or lowered those grades in a way that does not accurately reflect the instructor's assessment of the best and the worst examinations.

Faculty is allowed relief points equal to six percent of the enrolment in the course. For example, in a class of 80, on the 65-90 scale, an instructor could raise five grades by one point each, or raise three grades by one point and lower one grade by two points, or any other combination of adjustments adding up to no more than five points in total; on the 4-point scale, the instructor could affect five grades by 1 each. Relief points are typically assigned before the Registrar matches exam numbers with student names.
"Extra Credit"

Since law school courses are graded on a strict curve (normalization), assignments referred to as "extra credit" are in effect "optional." Since students' grades are assigned in relation to the grades of other students, if one student completes an "extra credit" assignment, that "extra credit" may result in lowering another student's grade. If instructors refer to assignments as "extra credit," students should understand that these assignments are optional but that the grades of students who do not complete them may be lowered as a result.

3.18.2 Birla Institute of Technology and Science, Pilani Model

In Birla Institute of Technology and Science (BITS), Pilani Rajasthan at the time of admitting students in their courses the candidate's position in the merit list is based only on his/her aggregate marks after it has been normalised by a process known as normalisation. The terms 'aggregate' and 'normalisation' are described below:

**Aggregate:** are the total marks of all subjects considered by the Board/University for computing result. The candidate's position in the merit list will be based only on the aggregate after it has been normalised by a process known as normalisation described later in this part. For this purpose the aggregate must contain the required subjects, namely, Physics, Chemistry and Mathematics.
Normalisation: To bring all candidates from the different examining authorities/boards on the same scale of comparison and to create a merit list in linear order, the Institute has been practising a system known as normalisation. It basically tries to find the relative displacement of a candidate from the candidate who stood first in the examination of the Board from which the candidate under review has passed. As such the first rank student of each board is considered to have obtained 100% marks and the aggregate marks of all other students from that board are normalised with reference to the aggregate marks obtained by the first rank student of that board.

For example, if the aggregate marks of the first rank student of a board is 94%, his normalised aggregate is 100% and if a student from the same Board has obtained 88% marks, then his normalised percentage will be \( \frac{88 \times 100}{94} = 93.61\% \).

3.18.3 Development of the Student Evaluation Standards

In USA a paper was drafted discussing the need for Standards for Student Evaluation. Currently the standards are a work in progress. While wider input has been sought it is important to note that these standards are in their draft form, the Joint Committee or any of its sponsoring organizations has NOT approved these standards.
During last few years the Joint Committee has and continues to follow its established Operating Procedures to review and revise the current draft materials. As input is gained at each point, the draft standards will be revised based on input received.

National and International Review Panels: The draft standards were subjected to national and international review.

Field-Tests: The draft standards were field tested in Class 12 and higher education classrooms.

National Public Hearings: The Joint Committee will hold open forums at meetings of each of the sponsoring organizations to gain additional feedback from members of these organizations.

Finalization of Standards: A task group will take input received to finalize the standards and the standards will be voted on, approved by the Joint Committee as a whole. Following Joint Committee approval, the Joint Committee shall also arrange for formal submittal of the final version of the standards to the American National Standards Institute (ANSI), in accordance with the ANSI Procedures.

Throughout the development process Joint Committee members will provide progress reports at the board meetings of their respective organizations, present papers and symposia, and offer training sessions in the use of these new standards.
Summary of the student evaluation standards (October 2001) is enclosed as an appendix 'A'.

Similar to student's evaluation standards there are Teachers' evaluation standards drafted. The summary of these evaluation standards is enclosed as an appendix 'B'.

3.18.4 Students' Evaluation of Educational Quality (SEEQ)\(^{35}\)

USA based Centre for Excellence in Learning and Teaching (CELT) has developed a computer based SEEQ system, which uses a standardised questionnaire based feedback drafted by Herbert W. Marsh. It is a 40 items questionnaire in machine scanable form. It is extremely useful for teachers teaching large classes. Most teachers use it in addition to the formal semester/yearly examination to evaluate their students.

Sternberg is currently developing the "Sternberg Multidimensional Abilities Test" to assess the three intellectual strengths. One wonders what future assessment instruments will evolve from this test, or in what ways or in what situations the test will be used for which it was not designed. Once again the mental measurement cycle begins. Many people treat intelligence as if it were a physical substance like hair colour or height. One may have "x" amount of it
"A study of grading systems in different universities for development of normalisation technique to streamline recruitment processes"

while someone else has "y" amount of it. Dr. Alexander Wesman of the Psychological Corporation wrote:

"We might better remember that it is no more reified than attributes like beauty, or speed, or honesty. There are objects, which are classified, as beautiful, there are performances, which may be characterized, as speedy; there are behaviours, which display honesty. Each of these is measurable, with greater or less objectivity. Because they can be measured, however, does not mean they are substances. We may agree with E.L. Thorndike that if something exists it can be measured; we need not accept the converse notion that if we can measure something it has existence as a substance". And yet the search goes on.

3.19 Some Grading Practices

After extensive search in locating if any standards are available in world over in terms of grading of the students either in vogue or in proposed forms, it is revealed that practices discussed in succeeding paragraphs are in vogue.

3.19.1 Grading Process

Barbara Gross Davis, University of California, Berkeley in her book “Tools for Teaching”, has described the grading process which is quite student friendly and addresses almost all the concerns that the students, their parents and academicians may have on this account. The process given out by Barbara
Gross Davis is quite generic in nature but addresses all the important issues and may thus be extremely useful. The details are enclosed as appendix ‘C’.

3.19.2 Grading Practices

The Graduate Studies Grading Practices falls within the general framework of the grading system of School of Graduate Studies (SGS), University of Toronto, Canada. This system has the features of Evaluation Procedures and Conditions for Final Standing, has been aptly described and is attached as appendix ‘D’.

3.19.3 In yet another paper titled TYPICAL GRADING PRACTICES three methods of grading have been commented upon namely:

(a) Relative Grading Methods (Grading on the Curve)
(b) Absolute Grading Method Based on Content
(c) Absolute Grading Methods Based on Fixed Scales

The paper compares all the above methods under four heads namely:

(a) Typical methods
(b) Benefits
(c) Drawbacks
(d) Sound Strategies

Brief on the comparisons on all the ibid three methods is at appendix ‘E’.
References:


5. 


6. SHAH, B. V. (1964), Social Challenge and College Students of Gujarat, MS University of Baroda, Baroda.


"A study of grading systems in different universities for development of normalisation technique to streamline recruitment processes"


11. Home Page : University Grant Commission, Description: Information on this government body, its activities, details of all funded universities and colleges. 

www.ugc.ac.in


13. Binet Alfred : Intelligence testing, Website 

www.pbs.org/wgbh/aso/databank/entries/dh05te.html


“A study of grading systems in different universities for development of normalisation technique to streamline recruitment processes”


www.uts.psu.edu/Classical_theory_frame.htm

30. Academic Testing: Item Response Theory Approach

www.uts.psu.edu/Item_Response_Theory_frame.htm

31. USC LAW SCHOOL Student Handbook: Grading System, New Grading System, and Grade Posting, Grade Normalization Policy,

lawweb.usc.edu/enrollment/HDBK/HDBKcon.html

32. BITS, Pilani (India) : Home Page www.bits-pilani.ac.in/

33. Development of The Student Evaluation Standards;

ec.wmich.edu/jointcomn/SES/SESOveryview.htm
"A study of grading systems in different universities for development of normalisation technique to streamline recruitment processes"


35. Centre For Excellence in Learning and Teaching, Student’ Evaluation of Educational Quality (SEEQ): www.psu.edu.celt/SEEQ.html

36. Barbara Gross Davis; GRADING PRACTICES. University of California, Berkeley. www.hcc.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/grading.htm


38. Pros and Cons of Typical Grading Practices Normal, Bell-Shaped Curve; depts.washington.edu/grading/plan/procon.htm