CHAPTER - II

REVIEW OF RELATED LITERATURE AND
EMPIRICAL RESEARCH STUDIES

The perusal of research literature in the area of examination, evaluation and assessment system, i.e. in the present area of investigation shows that the main work in this field has been done during the last four decades in foreign countries. In India, according to the Indian Education Commission (1964-66) examination and evaluation constitute one of those areas in education about which one can say, that the problem is known, otherwise the number of research studies are very few. A number of institutions in India have undertaken studies in some form or the other in the area of examinations, evaluation and assessment system for the last thirty years.

By doing so the investigators tried to identify trends in related researches and focus on the findings and conclusions of the researches that have some bearings direct or indirect on the present investigation.

Morty Geoggoj (1988) said that the review of literature promotes a greater understanding of the problem and its crucial aspects and ensures the avoidance of unnecessary duplication. It also provides comparative
data on the basis of which one can evaluate and interpret the significance of one's findings.

An extensive work of presenting reviews of Ph.D. theses completed in Education in India was done by Buch (1974; 1978; 1983; 1988). A good number of relevant research papers are published in several journals of education. It is rather difficult to take an account of each and every work done as the time and resources available are very limited. Hence, a few relevant researches are reviewed here. A brief and initial review and appraisal of the related studies having bearing upon the problem and some of those indirectly related are presented in the following pages.

Bokil conducted series of studies on SSC examinations. For example, some of the important studies are:
1) study of SSC examination results (1956), 2) impact of different media on the performance of students appearing for the examination of March 1955 and also on the percentages of failure (1956), 3) school-wise failure analysis of the SSC examination of March 1955 (size and location as factors) (1956), 4) investigation into the problem of relationship between the marks obtained by candidates in the various subjects at the preliminary examination held by the S.S.C. Examination

A few important conclusions from these studies were -

1) The percentage of the successful girls were higher than the boys. 2) The decline of examination efficiency with the advantage of age was clearly discernible for both boys and girls. 3) the use of respective regional languages help the students to secure higher marks except in mathematics. 4) The average ratios indicate a greater number of students failing in more subjects for rural schools than urban schools. 5) No close relationship between English and other language is observed at higher level. 6) The pupils did better in the S.S.C. examination as compared to their performance at the preliminary examination. 7) There was high correlation (0.8) between the two examination marks in aggregate total. 8) Most of the compulsory subjects showed low correlation than the optional subjects. 9) The results of the investigation did not support the conventional belief regarding the disparity in the performance of the candidates between urban, semi-urban and rural schools.
Buch et al. (1961) conducted a study on "Analysis of the marks in tutorials and annual examination". Main objective of the study was to analyse the marks obtained by the students in tutorials and the university examination and to know the effectiveness of the present examination system. Following were the main findings of the study: 1) Tutorial system of instruction had been utilised for assigning internal marks, thereby treating the tutorial system of instruction on par with the system of internal assessment. 2) In almost all cases, a tendency of liberal marking was observed. 3) There was a low correlation between two sets of scores, and in some cases, there was no correlation at all.

Lele et al. (1962) examined "variations in and relationship between student performance in paper examination and class record." The system of internal assessment had been introduced since 1957. According to this system 30 percent of marks were allotted for the internal assessment and remaining 70 percent marks were allotted for university examination. Main hypotheses of the study were: 1) The two procedures of evaluating student performance, paper examination and class record, are not related except by the error of chance. 2) The difference between the means of the two procedures of evaluating students' performance is nil except by error of chance.
The main findings of the study were: 1) The mean percent marks of the students in the two procedures, viz., paper examination and class record, varied significantly. 2) The paper examination mean percent was greater than the class record percent. 3) The relationship between the marks obtained through two procedures was significant.

Raina (1964) studied relationship between external examination marks and internal assessment of M.Ed. students. Major objectives of the study were - 1) To study the achievement of M.Ed. students in external examination and sessional work in different papers and dissertation. 2) To study the disparity between the two assessments in the theory papers. 3) To study the extent of the homogeneity between the examination marks, sessional marks and dissertation marks. 4) Contribution of sessional marks and dissertation marks in result and division.

All M.Ed. students of post-graduate colleges of Rajasthan University during 1959-63 were taken for the study. The following were the major findings. 1) There was no significant relationship between the external examination marks and the sessional marks. 2) The weaker students in the external benefitted more from sessional work. 3) Relationship between sessional work marks and dissertation marks with external marks held constant but
the negligible relationship was found between external examination and sessional work.

The objective of Taylor's (1964) study was to examine the reliability of the examiners. In this experiment double marking system was carried out for the candidates of college, all the subjects were involved, and every care was taken to make random arrangement of the answer scripts. Statistics like M, SD, Q, etc. were worked out. It was found that there were significant differences between examiners who produced very marked fluctuations in the classification of the candidates. For example, one examiner failed eight candidates, whereas other failed none. It was further found that a single examiner showed considerable changes in his standard of marking as he worked through the scripts. Different examiners showed large differences even in Standard Deviations.


Some of the major findings of the above mentioned
studies were: 1) The scaling tables provided a sufficient and satisfactory answer to the problem of mark adjustment. 2) These tables were useful for scaling sets of marks relating to the same question paper for which variations in the standard deviation were usually small. 3) All the tests clearly indicated the presence of persistence. 4) No significant difference was found between one subject and another. The effect was almost always positive. 5) High serial correlations appeared in at least one-third of the mark sheets. 6) The significant educational implication is that essay type examining, when properly conducted and analysed, can give more reliable results than is commonly supposed. 7) Nearly half the examiners were subject to large fluctuations of standard in the course of their marking. 8) One-third of the examiners exhibited large fluctuations in accuracy. 9) One-third were subjected to the 'persistence effect' by which the impression formed on one script was carried over to the next. 10) There were indications of a diurnal fluctuation in marking, and clear evidence of occasional periods when an examiner's judgement effectively ceased to operate. 11) The errors which had arisen from variations in the standard of marking were by no means negligible and their removal made a striking difference in the final results of the examination. 12) The scaling increased the pass
percentage and produced significant changes in the classification of the candidates and the order of merit.

An analytical study of answer books of a particular university examination at the graduate level was done by Chauhan (1967). In the first part of the study an analysis was done to explain the causes of failure. In the second part, the content analysis of some important issues related to the level of performance of candidates were examined. Two hundred and eighty-five answer books of one theory paper of the examination of the final year were taken for the study. Answer books were examined by the author himself, but the paper was set by a different examiner.

Some of the conclusions drawn by the researcher were:
1) Large proportion failed not because of not knowing the subject matter, but because of some internal factors like defective question paper, carelessness of examiners in evaluating answer books, evaluating by incapable examiners.
2) Hardly 60 percent of the total content presented by candidates was relevant and effective.
3) Expression of the candidates was generally poor.
4) Candidates generally made a wasteful use of paper.
5) A gap between the medium of instruction, the medium of study and of expression, and the language in which the question paper was set for examination created many problems and contributed significantly to the failure in examinations.
Bose (1967) conducted a study on "the existing system of examination and measures for improving upon it." The study investigated prevailing system of examination and suggested short term remedial measures. This was an opinion survey. A questionnaire was circulated to teaching institutions and persons interested in education. The questionnaire included items on syllabus, the setting and the moderation of question papers, conduct of examination, college tests, examination for external students, tutorial work, number of examinations and reports from colleges to the university.

The findings of the study were: 1) College teachers and principals branded the existing syllabi as too heavy. Most of the principals and college teachers wanted to change the syllabus. 2) A question paper should provide for double the number of questions to be answered and the questions should be distributed uniformly throughout the syllabus. 3) The course material should be divided into groups. 4) The university should undertake all the examinations in respect of collegiate education. 5) Opinions differed in respect of the method of teaching in colleges. 6) The necessity of tutorial/seminar work was not accepted in the same spirit by all.

To know the inter-examiner reliability of marks given by practical examiner under the conditions prevailing
in schools in Rajasthan, a study was carried out by Patel (1967). A series of three tryouts of practical question papers were conducted, different groups of pupils were tested in different subjects for different tryouts. Two types of practical skills were tested—instrumental skill and problem solving skills. These skills were tested in terms of process of performance and product of performance. Results showed that there was a very high agreement among the examiners in science subjects. There was a systematic increase in agreement from tryout to tryout in the biology practical examination. This clearly indicated the corresponding increase in the competence of examiners by means of training.

An investigation into the present system of tests and examinations—both internal and external—in the secondary schools of Madras state was done by Rao (1968). Main purpose of the study was to investigate the present systems of tests and examination in the secondary schools of Madras state. Two hundred twenty-one schools were selected for the study. For studying the internal assessment six schools were selected at random from the total sample. Results showed that in many cases the question papers were not properly balanced as far as the difficulty values of the items were concerned. They failed even to discriminate pupils of high from low abilities. The values
of correlation coefficients between the external and internal assessments were found to be greater in the language subjects as compared to other subjects. It was found that 15.10 percent of the students who were expected to pass were declared unsuccessful and 4.10 percent of the students who were likely to fail were declared successful in the public examination.

In an extensive study Jhavery and Patel (1968) tested following hypotheses: 1) There is significantly higher intra examiner reliability in case of marking an essay with a scheme than marking without the scheme. 2) The intra examiner reliability is significantly higher in case of well defined essays than the traditional essays. To test these assumptions, with the help of a workshop organised under the joint auspices of the NCERT and the SSCE Board of Gujrat state, different types of essays were prepared, and listing of different abilities required for effective essay writing was done. Findings revealed, that 1) The use of marking scheme significantly increased the mean assessment. 2) The use of marking scheme did not increase any type of reliability. 3) The reliability of the assessment of well-defined essays was found to be greater than that of traditional essays.

To examine association between internal and external
assessment, Kamat (1968) utilised data of 400 Ss. The data were taken from the records of university. Results of arts and science examinations from 1959 to 1962 were compared, the distribution of internal and external assessments in each subject were compared and the correlation coefficient between those two assessments in each subject were found out.

The study revealed that coefficient of correlation between internal and external assessments was not high, it was smaller in the arts than the science subject, it was smaller in the colleges in mofussil areas than the colleges in Poona city, smaller in newer colleges than older colleges, the average internal assessment was also generally higher than the average external assessment and again the difference was greater in the mofussil colleges than in Poona colleges and greater still in the newer mofussil colleges. The dispersion of the internal assessment was generally greater specially in arts subjects than the dispersion of the external assessment.

As a result of factor analysis of the examination marks, three distinct factors were identified.

Harper (1970) conducted several studies. In one of his studies he made attempt to find out the reliability of examinations in India. In two different experimental
conditions, 4010 answer books from four different high
school subjects were marked for the second time by 130
examiners. The marking reliability was found by two
experiments, first experiment was called the 'Ninety
marking ten' experiment and the other experiment was
called the 'four thousand re-examined' study.

The study revealed that in comparison to the results
from other countries (France, England, USA), Indian
examiners seemed to be slightly more reliable. This
statement applied only to 'untrained' examiners. Studies
abroad indicate that special training can greatly increase
the reliability of marking. Thus the examiners as such
are not unreliable but it is this particular system of
examination being essay type, types of questions, scaling,
inadequate instructions and training for examiners, that
are responsible for the low reliability.

Bennur and Mistra (1970) presented a paper in a
seminar on 'Examinations in Higher Education'. They
reported correlation coefficients between internal and
external assessments for two years as 0.41 and 0.56 for
B.Ed. examination, 0.04 for second year engineering
mathematics and-0.38 (not significant) for M.A./M.Sc./
Mathematics students.

The seminar suggested to concentrate upon issues
related to redesigning of question papers, use of marks, grades and even percentiles as bases of awards, aids, strengthening examination wings, and introducing semester system of examination. Finally, the seminar called upon the U.G.C. and the inter-university Board to organise a well staffed unit at the National level for research into problems of examination.

Mishra carried out a number of studies on examination system in India (1969, 1970a, 1970b, 1971). One of his studies tried to search, 1) whether the material covered by essay type tests could be made objective, and 2) if the merits and limitations of the essay type test be compared to the objective test. In each subject two parallel objective type tests and essay type tests were constructed. Following findings were found. 1) All the questions asked in traditional essay type tests could be covered by objective type items. 2) The coverage of course content in an objective test was wider than in an essay type test of the same duration of time. 3) Due to wrong selection of questions, students of higher ability might sometimes get lower marks than those of the students of lower ability in essay type tests. 4) The reliability of essay type tests was very low. 5) The predictive validity coefficient of objective test was higher than essay type test.
Reliability of external and internal marks of Vidarbha Board of Secondary Education examination was tested by Deshpande (1972). The study was undertaken 1) to find out the manner in which the reliability of the entire examination results is affected by the internal assessment system. 2) To evaluate the reliability of these two forms of assessment through statistical technique. 3) To consider the possibility of combining different assessments to produce a more acceptable composite for examination success. A random sample of twenty schools was selected from the Vidarbha region for the study. The following were the major findings. 1) Correlation between external examination marks and internal assessment had been invariably higher than the correlation between the objective test scores and either of the two forms of traditional examination. 2) The external examination appeared to be closer in its assessment to the objective assessment. Internal assessment was closer to the assessment by objective type test. 3) The correlation between external examination and internal examination fluctuated considerably from school to school within the same subject. 4) A composite score of external and internal assessment did not show any significant gain. 5) The internal assessment showed deflected range of marks as compared with the external examination marks. 6) In home examination subjects internal assessment had been most liberal.
7) Inter examiner fluctuations in marking within a subject appeared to be very high. 8) The third home examination and others thereafter showed consistently higher correlation with the external examination.

Shah (1972) studied the internal evaluation system in colleges, in different faculties, the scheme of distribution of marks allotted to the internal evaluation, weightage given to the different activities, the records of internal marking, difficulties in college assessment, reactions regarding prevailing system from principals, teachers and students, comparison between university examination, college tests and tutorial work, and other aspect regarding evaluation system. Some of the valuable conclusions of the study were, 1) The general pattern of internal evaluation is identical in arts, commerce and science colleges. 2) There is a trend of decrease in the weightage given to the internal evaluation. 3) Internal evaluation has increased the administrative and clerical work of the principal. 4) The teachers and the principals are not in favour of semester system. 5) The means of internal evaluation are reported to be significantly higher than university examination marks. 6) There is a significant positive correlation between internal assessment marks and university examination marks. 7) College tests are found to be significantly higher than
university examination marks. 8) Tutorial marks are found to be significantly higher than college test marks. 9) Correlation between the tutorial marks and the college test marks are not found significant.

To ascertain the factors influencing examination results and to suggest certain measures for their improvement Nath conducted a study (1979). Out of seven thousand candidates, a representative sample of 601 Ss was selected. Out of eighty-nine colleges forty-three were selected. He found that - 1) the highest number of colleges in the below average category was in the subjects of economics, English and geography. 2) The poor standards in results were mainly due to two factors - a) poor quality of students admitted, b) high student-teacher ratio.

In a paper presented at the seminar on examination reforms organised by the University of Bombay in August 1974, Kanekar (1974) presented a critical study of internal assessment as implemented at the post-graduate level in the University of Bombay. He stated that the university deserved to be congratulated for having taken a bold step towards examination reforms. It also deserved to be criticized for the fact that their internal assessment scheme was at best a half-hearted reform and at worst the travesty of a reform.
The present system appears to be hastily cooked up and the way it undergoes modifications from time to time displays a penchant for expediency rather than a concern for academic principles. Decision about university matters are influenced by people who really are not directly concerned with those matters, and related teachers are ignored. Moreover assessment is made by a teacher who does not teach a paper and breakdown of marks is also not proper (60 percent Internal + 40 percent External).

The present Internal assessment scheme and the way it is being implemented are saddening as well as sobering events in our academic life. Some students are unhappy because internal assessment requires more and better work. Some teachers are unhappy because of the extra work involved without additional remuneration. One sometimes hears comments on the effect that internal assessment has failed. But if there has been any failure at all, it is we who have failed internal assessment.

Problems in assessment and examination system are not characteristics of higher education only, but even at primary level similar problems are being faced. Tewari (1975) in a study searched the problems faced by teachers in evaluating the upper-primary children. Results showed that more questions were based on the knowledge aspect.
The percentage of questions on understanding and skill was very low. 2) In class work, essay type questions were mainly used. In tests, half-yearly and annual examinations, the position was reverse. 3) In multiple choice questions, most of the distracters did not function well but the stems were alright on the whole. 4) The teachers didn't know about the preparation of the blue-print. 5) The teachers didn't have mastery over their subject. 6) Teachers were not trained in evaluation techniques. 7) Teachers were ignorant about the objectives of unit-wise teaching.

Attitudes of students, their teachers and guardians towards the academic, evaluative and administrative aspects of the existing system of university examination was studied by Sinha (1977). Five hundred sixty, Ranchi University students were selected for the study. A Likert type questionnaire was constructed and used. Kallys's technique was also used. Some of the major conclusions were, 1) The existing examination system had both merits and demerits. 2) According to students, teachers and guardians, it had more demerits than merits. 3) A very high percentage of them had agreed with various suggestions for improvement in the existing examination system. 4) Most students, teachers and guardians were dissatisfied with the existing examination system and wanted improvement in it.
A study of relationship between marks obtained by students in theory and practical examination in science at S.S.C. examination was carried out by Sali and Umathe (1977). The study aimed at investigating the relationship between the marks obtained by the students in theory and practical examinations in science at the S.S.C. examination 1977 and suggesting a reliable marking scheme in practical examination in science. The hypothesis of no significant relationship between the theory and practical examination marks was formulated.

The sample comprised of 800 pupils selected on a random basis out of about 80,000 students who appeared in the examination in the Vidarbha region of Maharashtra.

The major findings of the study were: 1) Out of maximum possible score of 120, the mean score of a candidate in science theory was 38.51 (32.10%) whereas the mean score of a candidate in science practicals was 21.61 (72%) out of the maximum possible score of thirty. 2) High percentage of marks were scored in practicals. 3) The results indicated the non-discriminatory nature of examination between good and poor students. 4) Coefficient of correlation between theory and practicals was 0.34. 4) The majority of the students (75% above) scored less than 40 percent marks in theory whereas above 80
percent students scored more than 60 percent marks in practicals.

Palsane and Khedkar (1977) conducted a study on 'Internal assessment - Objective test and Final examination'. Main objectives of the study were - 1) To compare the internal marks with the external marks. 2) To study the relationship between internal marks and external marks at the B.Sc. examination. 3) To study the relationship between the internal marks and the external marks with M.Sc. semester I. 4) To study the relationship between the two components of the B.Sc. marks with the M.Sc. semester I. 5) To study the relationship between the two components of B.Sc. marks and two components of the M.Sc. semester I marks on the one hand and performance on the objective test on the other. 6) To examine the consistency of findings over the different sub samples.

Data were available from records for the variables of internal, external and total marks at the B.Sc. and at the M.Sc. semester I examination. An objective test consisting of two parts (36 items and 32 items) was administered to the students just as they were admitted to the M.Sc. course.

Some of the major observations and suggestions were, 1) The internal marking seems to be more liberal than the
external, though there is always a possibility of genuinely superior performance on the internal assessments. The divergence in the internal and external assessments varies from one sample to another. 2) The internal and external assessments at B.Sc. are negatively correlated with each other in these limited selective samples. The same show positive relationship at the M.Sc. semester I. 3) When related to the objective tests, the M.Sc. semester I assessments show higher positive relationships than the B.Sc. marks. Of the M.Sc. assessments, the external assessments are more consistent with the objective assessments than the internal marks.

Reddi (1977) conducted a study on "A comparative study of the attitude of post-graduate students of two institutions towards internal assessment". An attitude scale developed to measure the attitude of students towards internal assessments was administered to a group of 450 students out of which 240 students belonged to a university college where internal assessment is being tried, and the rest belonged to another university college where there is an idea of introducing internal assessment, at present the traditional system of examination being in vogue. The scores of the students were analysed using analysis of variance.
It was observed that first year students had a more favourable attitude towards internal assessment than the second year's. There was no significant difference between the attitude scores of arts and science students. This was true in both colleges. Men students of the university where internal assessment is being tried had a more favourable attitude towards the system compared to the women students, while in other university where the traditional system of examination still continues, women students favoured internal assessment more than the male students did. On the whole there was no significant difference between the mean scores of the two institutions. Seventy three percent students of the first university wanted the system to be continued. But only 16 percent of them were for complete internal assessment. Fifty two percent students of the second university wanted to change over to internal assessment, among whom only 8 percent wanted complete internal assessment. Many wanted partial assessment.

Bhat et al. (1978) conducted a comparative study of the grade table and direct grading method to test the consistency of grade distribution within the examinations and between the examiners using the two methods of evaluations.
The sample consisted of fifty answer scripts from among 397 scripts relating to the paper on administration of M.Com. examination. Four examiners were selected from the list of examiners approved by the university. A questionnaire was used to elicit opinions of the examiners regarding the practicability and suitability of the methods. Chi-square test was applied to test the association between the two sets of grades.

Three out of the four examiners were consistent within themselves in respect of the grades under grade table method. Each of the four examiners showed consistency within themselves in grading the scripts under direct grading methods. The examiners were consistent between themselves when the scripts were ranked in order of merit under both the methods. There was no agreement between the examiners regarding the distribution of grades by either method. There was more consistency in DGM evaluation than in GTM. Under GTM, the marks awarded by the four examiners ranged from zero to nine but the grade points awarded under DGM are zero to maximum point 7. The objectivity in valuation was more in GDM than in GTM evaluation.

To compare the performance of students regarding their internal assessment marks at traditional and semester
examinations, and to find out the relationship between internal assessment marks and final theory marks at traditional and semester examinations, Kenta and Sharma (1978) conducted a study.

One hundred seventy nine post-graduate students (81 from traditional and 98 from semester examinations) were taken for the study. Two batches under traditional system examinations and two batches passed under semester system of examinations were compared. From the findings of study following conclusions were drawn. 1) The range of mean sessional marks at semester examination was greater than the range of mean sessional marks at traditional examination. 2) The mean sessional marks was greater at semester examinations than traditional ones in almost all subjects. 3) The relationship between sessional marks and final theory marks at a semester examination was significant for all the four groups of subjects while it was not significant at traditional examination.

The merits of internal assessment were elaborated by them as follows. 1) This system inculcates habit of regular work among students. 2) Avoids guessing and elimination. 3) Teacher gets a clear idea of the strength and weakness of the students. 4) It removes unhealthy examination strain, stress, phobia and fever among the
students. 6) Multiplicity adds to reliability.  
7) Reduces undue emphasis on the final examination.  
8) Regular information of one's performance creates a healthy competitive spirit among students. 8) It helps the teachers to search remedy for any deficiency in learning on the part of the students.

Bhushan (1978) designed and conducted study for experimental verification of various methods of examination in history at the lower and higher stages. Major objectives of the study were: 1) To compare different types of examination. 2) To examine the reliability of scoring of examiners. 3) To study the significance of difference between internal and external examiners. 4) To find out the reliability of grades and works. 5) To study interrelationship between various types of examinations. 6) To assess overall discriminations of various types of examinations.

Major findings of the study were: 1) The internal examiner discriminated better among examiners in all these types of examination. 2) Wide divergencies were observed in marking standards at both levels in essay type, short answer and book type questions. 3) At higher stage, short answer type tests worked well, whereas at lower level essay type of examination worked well. 4) Evaluation through marking or grading had equal
importance at the lower stage but at the higher stage grading was better.

Scaling of some significant innovations in examination system was studied by Kaul (1979). Major objective of the study was to investigate the views of university teachers and students regarding the relative scale positions of significant innovations and to compare the two sets of ratings. Two hundred fifty post-graduate students and 80 teachers were selected randomly from various faculties of Post-Graduate Centre and Directorate of Correspondence Courses of Himachal Pradesh University. Major findings of the study were - 1) There was a good deal of similarity in the scale position of the innovations according to the views of the teachers and the students. 2) Certain innovation like introducing the grade system of ranking, essay type examination with objective type examination, semester system and supplementing external examination with periodic internal assessment, were in the top regions of the scale of the teachers and the students; collective opinion of both the groups was in favour of the grade system, internal assessment with external examination, semester system and objective type tests. 3) The teachers and students differed about supplementary written examinations with oral tests. 4) Innovations like open
book examination, use of computer, revaluation, spot valuation did not get importance from both, the teachers and taughts.

Main purpose of the study "A critical appraisal of some innovations for the improvement of examinations" by Verma (1980) was to make experimental validations of some innovations implemented in the examination system of universities. These innovations included marking and grading, internal and external assessment, improvement of question papers, open book versus traditional examinations, spot evaluation and revaluation. Researcher used grading system, five point scale, seven point scale and numerical scale of 101 points for evaluating the answer scripts. Some of the major findings of the study were –

1) There was no significant difference between the average assessment on five and seven point scale. 2) The internal assessment scores highly correlated with the external assessment scores where the number of students was small. 3) Inter examiners reliability of the scores awarded by different examiners improved with the improvement of the question paper. 4) Correlation coefficient between the scientific aptitude test and essay type test was low.

Association of Indian Universities Report (1980) submitted to the Ministry of Education, Government of India, in 1980, attempted to quantify the status and the
manifestation of internal assessment as currently practised by different universities. It was mentioned in the report that questionnaires were sent to 93 universities. It was revealed that 7 universities did not have any system of internal assessment, and their mode of assessment was fully external. There were 14 universities where the mode of assessment was fully internal and in rest of the universities the mode of assessment was partially internal. The components of internal assessment in all the universities were almost the same and they were quiz (announced/unannounced); short answer/long answer; essay type questions; class/home assignments; guided individual/group projects; rating; observation, check-list; laboratory/practical field work; seminars; group discussions and dissertation, etc.

The objectives of the system in practice were to test those skills/abilities which could not be tested through an external end examination, to integrate teaching and evaluation and also to make the teachers of the university/institutions identify the abilities/skills they wanted to develop in their students in their own subjects and then allocate weightage to the selected abilities/skills.
It was revealed from the information that as many as 23 universities/institutions introduced the system along with semester/trimester system, 21 universities/institutions along with grading system, 10 universities/institutions along with question banking and 19 universities along with restructuring of courses. There were only 4 universities and one institution which introduced the system in isolation on its own.

It was found that the teachers of 12 universities were of the opinion that the system is superior to the external system. As against this, the teachers of 7 universities were of the opinion that the system is inferior to the external system. Similarly, the students of 11 universities were of the opinion that the system is superior to external system and the opinion of students of 6 universities was that it is inferior to the external system.

Out of 93 universities, internal assessment has been introduced in 74 universities in one or other form. Out of these 74 universities, 13 universities gave a weightage of 20 percent, 6 universities of 25 percent, 7 universities of 30 percent and 9 universities of 40 percent. There were around 26 universities wherein the weightage of internal assessment marks varied from 10-15 percent and
there were around 14 universities wherein the range varied from 60 to 100 percent.

In almost all the universities, there was a machinery for looking into the grievances of students regarding the assessment of their performance in sessional work.

Rasool and Sharma (1981) conducted a comparative study of internal and external awards at the post-graduate level. Specific aims of study were - 1) To examine the characteristics of distribution of the scores awarded by the external and internal examiners. 2) To find out the degree of relationship between external and internal marks. 3) To find out the effect of internal marks on the boosting of the overall result of students. The study was conducted on 218 post-graduate students and 13 post-graduate departments of Jammu University.

Major findings of the study were: 1) The scores in most of the papers showed deviation from the normal distribution. 2) The range of the external marks was definitely more than that of the internal marks. 3) Most of the coefficients of correlation appeared to be positive. 4) The marks awarded under the internal assessment helped the students in raising their aggregate percentage of marks.
A study of the use of assignment activities and projects in teacher trainings was done by Gupta (1982). Main purpose of the study was to find out the most effective and useful method of the best training and education to the prospective teachers. Some of the useful recommendations by the researcher are as follows:

1) The training college should be residential. 2) There should be a variety and spontaneity of various hobbies and social activities. 3) Teachers should be trained before they actually start the actual teaching work. 4) Curricular research and co-curricular activities should be established. 5) Training of the future teachers adequately, for democratic co-operative living is essential. 7) Demonstration schools, which will serve as laboratories to the training colleges should be established. 8) Tutorial periods should be developed in almost all the subjects in the training colleges.

Main purpose of a study was to evaluate the effectiveness of some of the innovative methods in comparison with the traditional lecture method and to see, whether these new methods could help the learner better and to show higher achievement. This experimental research was conducted in the colleges of education in Uttar Pradesh. Main conclusions of the study were: 1) Discussion, symposium and supervised study method have been proved effective
than the lecture method. 2) Discussion method proves to be very helpful to the lower intelligent group.
3) Symposium method has given better results and proved to be definitely more useful to the average group.
4) Seminar method results are also useful. 5) Workshop method has also proved to be definitely superior to the lecture method. 6) Assignment method has not proved its effectiveness. 7) Supervised study method has also given much more better results.

Pramila Dabir (1984) carried out a study entitled "A critical analysis of the marks at the B.Ed. examination to study the trends and reliabilities of the assessment". The study was advanced on the basis of following hypotheses:
1) There is a low correlation between marks in theory examination and marks in the practical examination.
2) Performance of pupil teachers in practical subjects is better than that in theoretical subjects. 3) The internal assessment of pupil teachers in government colleges is less than that of pupil teachers in private colleges. 4) Private colleges are more liberal in giving internal marks than other colleges. The purpose of the study was to critically appraise the B.Ed. course of Nagpur University.

The marks in the B.Ed. examination for five consecutive years were collected from the records of Nagpur
University. An interview schedule was used as a tool to collect information from principals, lecturers and pupil teachers about the practical aspect of this examination.

Some of the major findings were: 1) The general percentage of passes was more than 90 for all the years. 2) There was an increasing tendency of passing and a diminishing tendency of failing. 3) The percentage of successful candidates in theory was remarkably higher in government colleges for all the five years. 4) The ranking order was, government colleges as first, private colleges as second, and university colleges as third. 5) There was wide disparity between the marks in the theory examination and the practical examination. 6) There were instances of students in private colleges scoring as high as 90 percent to 99 percent marks in the internal assessment.

A speech was delivered in the symposium on 'why examination reform?' conducted by Amravati University, Amravati by Natarajan on the topic, 'Basic Issues in Examination Reform'. The paper undertook, 1) Present syllabus and the semester type of syllabus, the methods of syllabus revision. 2) Frequency/Number of examinations and their implications. 3) Importance of instructional objectives. 4) Design of question papers/Restructuring
pattern of question papers. 5) Type of questions. 6) Type of optional questions. 7) Language of examinations. 8) Paper setter and examiners. 9) Question banking. 10) Internal assessment. 11) Grading system. 12) Preparation of teachers and students. 13) Processing of examination results.

Major suggestions were - 1) Syllabus must be prepared carefully and scientifically with specific clarity. 2) Semester examination at the end of first, third, fifth, etc. may be undertaken by colleges. A better alternative is to get the first two semesters examined by colleges and the next year (second year) semester examinations be conducted by university. 3) If objectives are to serve as guides to teaching and guides to evaluation, the way they are stated is of great importance. 4) The procedure for designing a question paper is tremendously important, it requires clear conception of objectives. 5) Question banking workshops for different faculties can be conducted at district levels. 6) Internal assessment could be started as pilot project. 7) Internal marks should be reported, along with examination marks, but the two should not be combined. 8) Grading system be introduced at graduation and post-graduation levels. 9) The university papers and published booklets, to be distributed to all teachers in the colleges, giving them essential information on the intended changes in examination system.
An investigation was designed to make a critical study of the system of examination in Kanpur University with a view of suggesting improvements.

The data were collected through study of the university records, and views of people were collected with the help of the questionnaire.

Main findings of the study were: 1) Internal assessment should be introduced. 2) Marks obtained in the internal assessment and the external assessment should be added up. 3) Central evaluation would be helpful. 4) The cases of unfair means should be decided at the time of examination. 5) The roll numbers on the answer books should be changed into code numbers. 7) The use of unfair means should be made a cognizable offence. 8) The publication of guess papers and guide books should be banned. The investigator, Kushwaha (1985), further suggested that internal assessment should be done by the subject teachers concerned.

A study of the perceptions of the school community in the city of Madras about reforming of the present examination system of the higher secondary school stage in Tamil Nadu was conducted by Jesudasan (1986). Main objectives of the study were: 1) To measure perceptions of the teaching community of higher secondary schools
in Madras city. 2) To examine critically about the academic and administrative components and issues related to the introduction of the internal assessment scheme and question bank as reforms in the examination system. 3) To study the relationship with their perceptions about the introduction of an internal assessment scheme and a question bank at the higher secondary school stage.

The tool for measuring perceptions, known as the Examination Reform Description Questionnaire was standardized. Reliability of the questionnaire was determined by split-half method and was found to be 0.97. The study was carried out on a sample of 500 teachers.

Major findings of the study were: 1) Eleven factors were identified. 2) Out of six independent variables, three variables showed significant difference. These were type of management, age and experience. 3) There was no significant difference between the perceptions of men and women teachers. 4) There was no significant mean difference between perceptions of science and humanities teachers. 5) There was no significant mean difference between perceptions of the post-graduate and inducted teachers. 6) In the validity factor, the mean score of non-government school teachers was higher than the government school teachers. 7) With regard to the difficulty factor, women teachers had a higher mean score than men.
teachers, while on the teacher student cordiality factor, men teachers had a higher mean perception score than women teachers. 8) In the validity factor, teachers above 40 years of age had a significantly higher mean score than teachers below 40 years of age. Teachers having above 15 years of experience had a significantly higher mean score on the validity factor than teachers having less than 15 years of experience. 9) In the facility factor, the teachers of humanities had a significantly higher mean score than the teachers of sciences.

Caroline Nagilian Kim (1990) measured attitude of post-graduate students towards internal and external assessment system. From findings of the study it was concluded that the students favour internal assessment as they consider over in general that through internal assessment they can overcome the fear of examination and the attitude scale was significantly reliable, where correlation was equal to 0.41.

On the whole, through the analysis of the data, it was concluded that the internal assessment was favoured by the students due to its merits over the external examination system.

Foreign Studies:

Wiseman (1956) conducted a study on 30 minute
compositions written by a group of 173 students, graded by four examiners. Correlation between the aggregate marks on the first and second essays was 0.89 and thereafter correlation for the same range of ability was 0.92. This implied that the total reliability of the essay test was high for the pooled marks of four readers.

Penfold (1956) explored the area of variability in marking essay test. Five minute essays written by 16,000 candidates were marked by impression by sixteen examiners, each examiner marking one thousand scripts. After an interval of some time 165 scripts selected at random from the whole batch were all re-marked by fifteen of the same examiners. The results indicated that there was significantly high variation between different examiners, standards of marking and also between the marks of some examiners on two different occasions of marking. Pidgeon and Yales (1957) and Valin (1961) also took up similar studies on the reliability of examiners' marks at different times.

Kloin and Hart (1968) investigated the change and systematic factors affecting the essay grades by getting the scripts of other 1500 students of low marked by seventeen professors. They found that students tended to be consistent in their performance in different essay questions used in the low school. The agreement in grading among the professors was 0.76.
Comer (1981) conducted a critical study of educational evaluation, theory and practice. This researcher argues that evaluation process is a moral activity of the examiner. In the final analysis, the professional evaluator can be seen as a moral agent and not simply a provider of information. Evaluation is essentially a technical activity characterized by fitting means to pre-identified ends. Language is the main instrument of the evaluation process.

Van Allen and George Howard (1981) carried out an analysis of the relationship between student evaluation of faculty and student faculty educator's attitude, similarity and selected variables. The North Carolina Community College System was chosen as setting for the explanatory story. Four hundred eighty eight students and forty faculty members were selected for the study. Some of the major findings were: 1) The analysis of data determined that student faculty attitude similarity was significantly related to student evaluation of faculty. 2) Variation in student evaluation of faculty was significantly (PR = 0.0225) influenced by educational attitude. 3) Significant relationship between the class evaluation of the teacher and similarity in faculty class attitude was found.
Students' attitude towards grading practices was assessed by Murray (1981). The researcher conducted this study with an inspiration that the issue of college grading practices is frequently debated among college administrators, instructors and students. That the issue is an important one is clear when one recognized that grades directly affect the career choices of many students. Unfortunately, through wrong practices regarding college grading, and the inexperience of most college administrators and instructors in the area of test development and educational measurement, considerable confusion exists about proper grading practices.

The research described in this study was designed to respond to several shortcomings of many earlier studies. Earlier researches have seldom considered the dimension of "Instructional setting" in their work. Secondly, too often researches have compared one grading system with another without regard for the intended use of grades in a particular instructional setting. Finally, it is important to consider both faculty and student views.

The first part of this study was conducted at the University of Massachusetts in 1976 by administering two similar instruments designed to measure faculty and student attitude towards use of grading in several
instructional settings, and toward the appropriateness of a variety of commonly used grading systems for accomplishing intended uses of grades. The second part was conducted in 1981 by administering a slightly modified version of the 1976 faculty questionnaire to the faculty in the school of education. The third part of this study was carried out to address the changes after a five year elapse of time.

Specially five questions were addressed:
1) What is the importance of various possible uses of grades in different instructional settings?
2) What is the acceptability of five common grading systems for accomplishing various common uses of grades?
3) What are the effects of the five common grading systems on a variety of course outcomes?
4) Which of a multitude of course outcomes are thought to be important in courses?
5) Which student factors should affect grades in different instructional settings?

The major findings of the study were: 1) Instructional setting had no effect on the ratings of importance of twelve common uses of grades. 2) Five uses of grades were rated as significantly more important than the others studied:
1) Inform others about student performance.
2) Provide students with feedback on their course.
3) Motivate students to do good work in the course.
4) Provide instructors with information about student progress.
5) Improve a student's ability to critically assess his/her own work.

These five uses remained significantly more important according to the faculty ratings on the 1981 revaluation. Three courses outcomes were viewed by faculty and students as more important than other: 1) Maximize students' enthusiasm. ii) Maximize student performance. iii) Maximize amount of learning.

From this study the researcher found that over 90 percent of faculty and students in the school of education found criterion reference grading to be 'at least minimally acceptable'. With respect to the most important course outcomes, again, criterion referenced grading was rated to be best.

Patton and Garner (1982) conducted a study on evaluation of the secondary teacher education programme as perceived by student teachers at the University of Alabama. Purpose of the study was to identify existing problems or justify current conditions and practices of
of the secondary teacher education programme and of the programme evaluation procedures. Some suggestions have been given about secondary education programme and evaluation procedures.

Glen (1983) conducted a study on 'A comparison of summative and formative teacher evaluation at a rural Missouri High School.' The problem was to determine which method of teacher evaluation is most valuable to the teacher as a professional. The evaluation procedure was discussed with the teachers at a post evaluation conference. The findings of this study indicate that the teachers prefer formative evaluation as opposed to summative evaluation. The teachers felt that the development of objectives enabled them to have input into the evaluation process and student awareness of objective made learning easier and increased accountability.

Day (1983) carried out an analysis of evaluation techniques and procedures implemented in selected Chicago area schools. Some objectives of the study were, to study most commonly used evaluation approaches, use of specific supervisory techniques, the rank value given to specific supervisory practices, the frequency of evaluation, evaluation criteria known to the teacher prior to evaluation, etc. The study sample consisted of 300
secondary Chicago area principals and assistant principals currently involved with teacher education. On the basis of findings some suggestions for the improvement of quality of the evaluation process were given.

Jeffrey (1983) conducted a study on 'the Design, development and field test of an evaluation framework for short term training programmes'. Main objective of the study was to identify an evaluation framework capable of assessing the impact of short term training programmes. Numerous methods were used to collect reaction, cognitive and behavioural data from multiple information sources. A meta evaluation was designed and conducted to assess the effectiveness of the field test evaluation. Some of the main findings were: 1) The most effective and efficient procedures were the end of week evaluation, final debriefing session and video-tape rating scale. 2) Discrepancies in evaluation results should be excepted when qualitative and quantitative data are gathered from variety of information sources using different evaluation procedures. 3) The evaluation framework is not useful for purpose of providing immediate format evaluation information to decision makers. Revised matrix reflected the results of the study.

This brief review of literature denotes the fact
there is much to be done in India for the improvement of both internal as well as external evaluation system in all faculties of education.