CHAPTER III

REVIEW OF RELATED LITERATURE

3.1.0 INTRODUCTION

3.2.0 RELEVANT RESEARCHES - INDIA

3.2.1 RELIABILITY

3.2.2 VALIDITY

3.2.3 SELF-EVALUATION

3.2.4 QUESTION PAPER

3.2.5 INTERNAL ASSESSMENT

3.3.0 RELEVANT RESEARCHES - ABROAD

3.3.1 RELIABILITY

3.3.2 VALIDITY

3.3.3 ORAL EXAMINATION

3.4.0 CONCLUSION
3.1.0 INTRODUCTION

It is needless to harbor the sad fact that in India, the country is under the regime of the examination in spite of the sincere and continuous reforms introduced from time to time by the ardent and enthusiastic educationists.

All grades, honours, degrees, posts, promotions, salaries, etc. are determined and secured by the examination results, from the beginning of a child's school career to his entry into a business or profession or into Government Service. The parent's ambition is that the child should pass. The teacher's success depends on his ability to prepare his students to pass. The passing of examinations has become a ruling motive throughout.

Whatever has been said by Herbert Rice can be found in Indian situation, and his observations are worth taking note of. Now it is time for education in India to be liberated from the tentacles of the Examination system.

Thus Examination has become a necessary evil in the life of an Indian student. A number of studies have been

conducted both at home and abroad on various aspects of the examination such as Internal Assessment, reliability, validity, self-evaluation, oral examination and setting up of question paper. A few such studies which are very relevant to the present study have been given below with their main findings with the hope that they would throw more light on the research on hand.

3.2.0 RELEVANT RESEARCHES - INDIA

In the forthcoming pages, some of the studies conducted in India with regard to Examination Reform are summarized. Some of the studies associated with the reliability of examinations are given below.

3.2.1 RELIABILITY

By reliability is meant the degree to which a test or examination measures what it really does measure, not necessarily what it purports to measure.

Some of the causes of the unreliability of examinations as listed by C.W. Valentine are given below.

1. The element of 'luck' in the type of questions a candidate happens to meet has too much weight.
2. A more serious ground for lack of confidence in examination awards is personal variability and unreliability of marking in the examiners themselves.
In subjects like English and History, it is more and more while less in elementary arithmetic. In the marking of essays, extraordinary variations occur between the marks of different examiners.

3. There is a further possibility of error in the influence of suggestion of a chief examiner, who in checking the marks of another is already aware of the marks previously given.

4. Another reason for suspecting the reliability of some examinations as a comparative estimate of the ability of candidates is that pupils come from different schools in which the standards of teaching (in different schools) may have been specially well taught have an advantage. 2

Thus the foregoing causes of the unreliability of examinations as pointed out by C.W. Valentine, in 1932 are rampant even in the present Examination system. Very little has been done to minimize, if not to eradicate, the causes of unreliability of examinations.

AwK. Gayen, et al, through their research on "Measurement of achievement in English" undertaken for the

N.C.R.T., New Delhi, concluded that

1. The present system of public examination is neither reliable nor valid because performance of students is assessed in single final examination by a large number of examiners on a set of questions mostly of the essay type, comparatively small in number and again students are given option to select alternatives.

2. If suitable arrangements are made in schools by appointing properly qualified teaching personnel, results of oral internal tests may also be incorporated in school records.

3. No examiners should be given too many answer scripts to assess.

The doctoral research by N.R. Sharma, on "An Analysis of Public Examination Results" indicated that

1. The reduction on curricular load led to significantly better average performance in all the individual subjects taken for study.

2. The rural candidates on an average got five or more marks less in every subject than his urban counterpart.


B.J. Jhaveri and B.C. Patel in their research on "A study of Inter and Intra Examination reliability in marking essays with and without using a Marking Scheme" found out that

1. The use of marking scheme did not increase any type of reliability.
2. The reliability of the assessment of well-defined essays was found to be greater than that of traditional essays.5

R. Bokil conducted his research on "Relationship between marks at Preliminary Examination of School and Secondary School Certificate Examination of the Board of Maharashtra" and concluded that there was a very high correlation (0.8) between the Preliminary and Secondary School Certificate Examinations.6

5. B.J. Jhaveri and B.C. Patel, A Study of Inter and Intra Examination Reliability in Marking essays with and without using the Marking Scheme, New Arts College, Sardar Patel University, Vallabh Vidyanagar, 1968.

L.N. Tluanga in his doctoral thesis on "Examination as a Mode of Management" has found out that

1. Marking standard was independent of marking spread.
2. It was found that an examiner's judgement when marking script is influenced by his impression of the proceeding script.7

Thus the foregoing pages deal with the reliability of examination and the following few pages are concerned with the validity of examination.

3.2.2. VALIDITY

By validity is meant the degree to which a test or examination measures what it purports to measure. The following are some of the major findings of validity of examinations.

J. Dennis Aras, for his doctoral degree, analysed the Secondary School Leaving Examination system in Kerala State and indicated that

1. The Kerala Test in Social Studies was invalid for the greatest part because it did not test most of the aims of the subject.

2. The test items were not determined by test outlines, difficulty indices, validity indices and practical purposes.

3. Recall and completion type of items were not at all correct.

4. Multiple choice and matching items offered opportunity for guessing.

5. Short-note items lacked objectivity and were essentially recall items.

6. Essay questions tested only information and little discriminatory power.

7. The methods of estimating reliability coefficients revealed that the scores were not reliable or consistent.

Therefore, the large scale elimination of students by test scores was unjust and unscientific.

8. Tests were inadequate and students needed teachers and administrators trained in counselling and guidance.

9. Reform is inevitable and is only a matter of time.

10. Teachers should adopt a system of total evaluation of the student. Therefore,

    a) Teachers should be given necessary training through in-service education

    b) Teacher education in Kerala in the future should
include courses on tests, measurements, guidance and counselling.

The research undertaken by M.N. Palsana, on "Predictive Potentiality of the Secondary School Certificate Examination Scores" revealed that the predictive value of the grand total marks at the S.S.C. Examination was found to be quite high.

H.J. Taylor, for the first time in India made systematic analysis of marks in a major examination namely Matriculation Examination conducted by the University of Gauhati and concluded that the examinations were unreliable and the validity of the examinations was poor. It is abundantly clear that the element of chance in the conventional examination is far greater than has previously been suspected on account of the unpredictable variations between the standards of different examiners. The fact that examiners work on a widely different standards, even under the most favourable conditions, has been known for many years.


After having reviewed some of the studies with regard to the validity of examination, it may be of some interest to know about another novel aspect of evaluation, namely self-evaluation.

2.2.3: SELF-EVALUATION

In her doctoral thesis on "Teaching Effectiveness and Student Evaluation of Teaching", Ethel Saline Jacob found out that teachers in Madras University area in general have a favourable attitude towards evaluation of teaching. Teachers estimate their teaching higher than their students do, whether the teacher's rating is the correct one or that of the students' is not the question. Student satisfaction is to be considered important as much as that of teachers have to take into account student satisfaction, while they evaluate themselves. Teachers are to be encouraged to do self-evaluation Vis-a-Vis student evaluation of teaching in order to become more cognizant of student feelings regarding their teaching.11

While evaluating knowledge of the subject, students give ratings based only on the acts manifested in the class. The teacher has to realize that it need not be

absolute rating of his mastery but only to the extent to which he manifests his subject mastery in the class as perceived by his students.

There is improvement in terms of student rating when student evaluation is fed back to teachers. Basically the improvements in the characteristics satisfy the learners and to that extent the teacher is successful.

While the foregoing page deals with self-evaluation, the following page deals with the different aspects of question papers.

3.2.4: QUESTION PAPER:

V. Bhola undertook a critical study of the effectiveness of the Matriculation Examination in Physics and Chemistry conducted by the Board of School Education, Haryana for his doctoral research and his findings are given below.

1. Alternative question papers set for the morning and evening sessions were not analogous and of same standard.

2. The format of the question paper was poor from the point of content, coverage and representation of the objectives of teaching Physics and Chemistry.

3. The question papers were not well-balanced in respect of attributes of discriminatory power,
difficulty value, reliability and validity of the question items. 12

B. Bhushan in his doctoral thesis on "Experimental verification of various methods of Examination in History at the Lower and Higher Standards" found that

1. There were wide divergencies in marking standards at both levels in essay type, short answer and open-book of examinations.

2. At the higher stage, short answer tests worked well, whereas at lower stage essay type of examinations worked well.

3. The abolition of option was favoured at both lower and higher stages. 13

The foregoing page deals with the different aspects of question paper and the page that follows deals with some of the studies on Internal Assessment.

3.2.5: INTERNAL ASSESSMENT:

Working for his doctoral degree on "Relationship


between External Examination Marks and Internal Assessment of M.Ed. students of Rajasthan University", T.N. Raina concluded that the findings of the study substantiate the hypothesis that there was no significant relationship between the external examination marks and the sessional work marks.

The weaker students in the external examinations benefit more with sessional work than the students who have better performance to their credit in the external examination.

The means of the external marks differ significantly from the means of the sessional work marks. The sessional marks may vary but very slightly.14

R.J. Roshiah undertook his doctoral research on "Desirability and Feasibility of Semester System". In his thesis submitted to the M.S. University of Baroda, he indicated that the students generally supported and welcomed the Semester System.15

K.T. Christopher in his doctoral thesis submitted to


the M.S. University of Baroda on "Perception of College Community in the city of Madras about the U.G.C. Action Programme" found out that the students and teachers welcomed the three dimensions (Question Bank, Internal Assessment and Grading System) of the U.G.C. Action Programme.16

3.3.0 RELEVANT RESEARCHES-ABROAD

Thus, the foregoing pages have dealt with some of the studies undertaken at home on the different aspects of Examination Reform.

3.3.1 RELIABILITY:

The following pages deal with the studies done abroad with some of the aspects of Examination Reform. At first, reliability of examination is reviewed.

As early as in 1923, reliability of old-type of examinations consisting of only essay questions was suspected. P.B. Ballard, selected a sample of 400 students, who were given two types of tests, one the old type of examination consisting of only essay questions and the other consisting of only objective type of questions. The researcher concluded,

When it comes to the question of reliability, the new examination wins easily.\textsuperscript{17} Thus, the desirability of the new examination consisting of objective type of questions has been well-established beyond any doubt as early as 1923. The saddening feature is that the examiners in India are still clinging to the essay type questions which has very low reliability. The coefficient of reliability of the new examination was found to be 0.905.

It was Hartog (1925), who, for the first time in the world, proved that two examiners never agreed among themselves.

G.M. Ruch, conducted his research on, "The comparative reliability of Five types of Objective Examination" and found out that

1. The reliability of the objective type per unit of working was very much greater than for the traditional essay type of examinations, reliabilities of 0.30 to 0.90, being possible for 20 minutes in comparison with 0.40 to 0.70 for 30 to 60 minutes written essay tests.

2. It is probably better to increase the length of objective examinations somewhat in order to eliminate chance successes rather than to give short examinations and depend upon statistical corrections.

3. Multiple-choice recognition examinations are easier (in the sense that higher average scores are concerned) than recall types.\textsuperscript{18}

C.W. Valentine, proved that results of the examinations were not at all reliable. Through his investigation on, "An Enquiry with special reference to the Entrance Examination to Secondary Schools, the School Certificate Examination" found that the top boy at the entrance test drops to 10th, four years later and the 3rd to the 18th. The bottom at the entrance, which only just succeeds in getting into the school, rises to nearly half-way up at school certificate Stage while No. 18 soars to 6th.\textsuperscript{19}

A report on an investigation carried out by a sub-committee of International Institute Examination Enquiry Committee, London, (1941), on the title "Marking of English Essays" concluded that two examiners never agree among themselves. What is considered as 'perfect' by one will not be considered as perfect by the other.\textsuperscript{20}


W.B. Robert Jackson and A. George Ferguson working on "Studies in the reliability of Tests" indicated that
1. Practice may lengthen or shorten the 'true' length of the test and hence affect the reliability of coefficient.
2. The length of time elapsing between tests seems to have little effect in the estimates of reliability.
3. The number of items composing a test is not a very efficient measure of the 'true' or 'effectual' length of a test.21

K. Lovel conducted his investigation on "Examinations and Marking" and arrived at the conclusion that the longer the examination and the greater the number of questions to be answered, the more reliable will be the score.22

S. Willmott Alan investigated on "The Reliability of Examinations at 16+". On the basis of the evidence presented, it was calculated that the C.S.E and G.C.E-0-level examinations were equally reliable and that the level of reliability attained was as high as might reasonably be expected.23

The investigation conducted by a team consisting of J.W. Britton, N.C. Martin and H. Rosen in the year 1976 is given below.24

The G.C.E composition scripts of representative sample of 500 boys and girls were marked by multiple marking methods and the results compared as described in validity and reliability with the official marking communicated to them by the G.C.E Examining Board.

1. **Validity** During the year leading up to the examination, candidates in the sample were given a monthly composition test in order to provide them with a broadly based criterion of their writing ability. The official marking and the experimental marking were compared in turn with the criterion to give an estimate of the validity of each marking.

2. **Reliability** Arrangements were made for each of the marking procedures, the official and the experimental, to be carried out a second and third time by independent markers or marking teams on a proportion of the scripts. The reliability of each marking method was estimated.

by directly comparing the three sets of results obtained in this way.

In the official marking, the correlation was 0.52. In the experimental marking, the correlation was 0.80. The figures clearly indicate that in this case of marking by individual examiners with very careful briefing and elaborate arrangements for moderation was in fact significantly less reliable than a multiple mark consisting of three rapid impression marks and a mark for mechanical accuracy.

The research conducted by N.J. Entwistle and J.D. Nisbert is summarized below. 25

687 essays were selected from Aberdeen. It was decided that five Markers A, B, C, D, and E should each give the individual essays a mark for their general impressions. The scale originally decided on was a six-point one ranging from 0 to 5. In addition, marker A and B each gave a mark for the mechanical accuracy of the essay (spelling, punctuation and paragraphing). Markers C and D each gave a mark for the essay's originality and content. Marker E gave a mark for vocabulary.

Product-moment correlation coefficients between markers were calculated for general impression marking on 687 essays. The values were found to be between 0.44 and 0.50. The agreement between two markers using mechanical accuracy as a criterion was indicated by a correlation of 0.45, while a value of 0.51 was found on originality and content.

Thus, the foregoing pages give a clear picture about the studies on the reliability of examination. The following pages are concerned with the validity of examination.

3.3.2: VALIDITY:

K.B.E. Philip and E.C. Rhodes verified the validity of the examinations. Their main finding is as follows.

Whereas the scripts had originally been all allotted the same moderate mark, they were allotted by these fifteen examiners 42 different marks varying from 21 to 70. It is also to be noted that the maximum is 96 and not 100.26

In his research on "Variability in results from New-type achievement rests", P. Pullias Earl indicated that

1. A test may be objective in the sense that all personal opinion is eliminated in scoring and still

---

fail to remove important personal elements from the evaluation of pupil achievement.

2. Measures of pupil achievement obtained from different informal objective tests may be expected to vary to a considerable extent. Thus, if a pupil takes Teacher A's test, his score, rank, and mark may be very different from what his scores, rank, and mark would have been, had he taken Teacher B's test. 27

In 1948, Allison Davis took the position that students in slum schools may well find it more rewarding to be considered academic failures than to be success in school.

To test the validity of Davis' observation, a study was conducted by H. Charles Josephson in a high school that serves some of Chicago's lower class neighbour. Although the group studied was small—only 106 students, the results of the study are surprising enough merit serious consideration.

In more than 75% of the cases, the grades, the student felt he deserved was the same or one point below his aspirations. And in more than 65% of the cases, the grade the student actually expected to receive was identical with or one point below the grade aspiration. These interpretations tend to dispute the value of a grading system such as the one followed by the school that was surveyed. Generally, the concern has been that grading is an inadequate

and often an arbitrary means of measuring learning. In addition, we seem able to conclude that grading as an evaluating process may actually serve as a deterrent to learning.

A specific grade serves as a magnet, but it is not a magnet, that draws the best out of the pupil. In fact, it may serve as a stimulus to failure for many students whose ability gives them no right to fail.

After reviewing about the validity aspect of Examination Reform of some of the studies in the field, the findings of a study on oral examination conducted in Germany are given below.

3.3.3: ORAL EXAMINATION:

K. Ingenkamp and B. Wolf in their thesis on "Research in Oral Secondary School Leaving Examinations in Germany" concluded their findings as follows:

Using an analysis of variance design examination sessions in literature and Geography were recorded on video tape and evaluated by 156 teachers of German Literature from 39 schools and by 81 Geography teachers from 24 schools. The results confirmed a remarkable between-rater variance.

The differences were significantly influenced by the speed of speaking, background information about part achievement.29

A good performance is rated better if it follows a weak performance, and a weak performance following a good one receives an even worse rating.

The strong influence of speed of speaking on marks reflects in our opinion most obviously the implicit personality theories of teachers.

3.4.0 CONCLUSION

Thus, an attempt has been made in this chapter to give a very brief account of some of the major findings on the various aspects of Examination namely, Reliability, Validity, Setting up of Question Papers, Internal Assessment and Oral Examination, both at home and abroad.

REFERENCES

I. BOOKS.


II UNPUBLISHED MATERIALS.


III PERIODICAL