An Analytical Study Of Some Aspects Of The Continuous Internal Evaluation System At The Tertiary Level

A Thesis

Submitted in Partial fulfilment of

the requirements of the degree of

DOCTOR OF PHILOSOPHY

BY

ANITA DUBEY

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI (Rajasthan)

July, 1981

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE,
PILANI, RAJASTHAN

CERTIFICATE

This is to certify that the thesis entitled 'An Analytical Study of Some Aspects of the Continuous Internal Evaluation System at the Tertiary Level' and submitted by (Mrs.) Anita Dubey Id. No. 76E93001 for award of Ph.D. degree of the Institute, embodies original work done by her under my supervision.

Signature in full of the supervisor

Name in Capital block letters V. V. MANDKE

Date July 10, 1981

Designation Professor & Dean Practice School Divisio

ACKNOWLEDGEMENTS

No words would be adequate for me to express my deep sense of gratitude to my supervisor Dr.V.V.Mandke, Professor of Electrical & Electronics Engineering and Dean, Practice School Division for his inspiring guidance throughout the course of this work. Without his help this work could never be completed.

It is my great pleasure to thank Prof.C.A.S.Naidu for his valuable suggestions and most helpful discussions. I am grateful to Mrs. Parimal Mandke for her personal interest in my work and all the help. I am thankful to all the members of 'Science and Society' seminar group and especially to the Group Coordinator Prof. H.C. Mehta for various fruitful discussions and suggestions. I also wish to thank Prof. S.C. Shukla for reading the manuscript and his useful comments.

I am grateful to all the faculty members of the Institute who were kind enough to supply the required data on examinations and also responded to my questionnaires. I received a lot of encouragement and help from all my teachers at BITS which made my task that much easier. In particular I wish to thank Prof. T.S.K.V. Iyer and Dr.V. Krishnamurthy, the Deputy Directors, for their interest in my work and encouragement.

I wish to express my sincere thanks to Dr. P. Dhyani who helped me with the computer programs as well as many other aspects of the work. I also wish to thank members of staff of IPC, in particular Mr. G.R. Saini for help in the compilation of data and to Mr. B.P. Singh and Mr. Ravindra Kumar for assistance in computer work.

I wish to thank Dean PS Division, Dean R&C Division, Dean F.D. III, Chief of IPC and Librarian for providing the necessary facilities and also to other members of staff of these divisions and units for their friendly and helpful attitude.

Thanks are due to Mr.H.S.Kulhari and Mr.M.K.Joshi for their typing work, and to Mr. K.N.Sharma for drawing the figures.

It is a pleasure to thank several of my friends who have given their generous help, encouragement and support. In particular, I wish to mention Mrs. Shree Devi S. Nidhi, Dr.(Mrs.) Amita Gupta, Dr.V.K.Dube, Mr.M.K.Kashiramka, Mr. S.K. Sharma, Mrs. Meera Banerjee, Miss Anuradha Mitra, Miss Sushma Mainra, Miss Neelima Mainra, Dr.(Mrs.)Meenakshi Agarawal and Dr.(Mrs.) Chendrakala Tripathi.

The award of research fellowships from the University Grants Commission, India and BITS are gratefully acknowledged.

I express my appreciation for the help given by my mother and other family members in particular Mrs. Sharad Tewary, Dr. V.K. Tewary, Anuranjita and specially to Kanupriya who helped in the checking and the organisation of the manuscript.

I wish to record my deep sense of gratitude to Dr.C.R.Mitra, Director of BITS for his interest in this work and his constant encouragement and help.

A very special word of thanks goes to my husband Anil who inspite of being several thousand miles away from me, constantly encouraged and supported me. Finally, I dedicate this thesis to my father who could not live long enough to see it completed but I could always feel his inspiring and strengthening presence with me.

Anita Dubey

(ANITA DUBEY)

ABSTRACT

The thesis has studied some aspects of the internal continuous evaluation system at the tertiary level. Towards this, the thesis has extensively drawn from the data-base as available through the educational operations at BITS, Pilani. To begin with the thesis has analysed the teachers' perception of the objectives of education as also the objectives, techniques and nature of examinations in the above context. Then for 83 examinations, consisting of 7 quizzes, 52 tests and 24 comprehensives, using the AOV approach, the thesis has studied the reliability co-efficients and the same have then been further critically analysed vis-a-vis parameters such as number of questions in an examination, number of students, examination duration, type of examination, level of the course, category of the course, course discipline, etc. The thesis has also investigated intercorrelations for as many as 303 evaluation component pairs, covering a spectrum of evaluation components extending beyond the traditional examinations of quizzes, tests and comprehensives. Further, the thesis has also studied the theme of course reliability. Next, the thesis has investigated the predictive validity for the student performance in Higher Secondary as also for the student performances in courses of 'Concepts in Science' and 'Modern Physics'. Towards this the data-base is provided by the 1976 input to BITS of 310 students and 1977 input of 316 students.

Further, the thesis has also investigated the theme of concurrent validity. This then is followed by study of the evaluation system as under the Practice School (PS) system of education. Using the data base as available from two sessions of the Nagda PS-II station, the thesis has investigated various aspects of PS evaluation such as: reliability of PS evaluation components of quiz and viva; examiner reliability for PS-II evaluation components of seminar, group-discussion, project report and observation; PS-II course reliability; inter-correlations between different PS-II evaluation components; content validity of the PS evaluation; etc. Finally, abstracting from the educational organisation as under the M.E.(Collaborative) programme, the thesis has discussed a model for a multiple objective assessment in a class-room based situation. Towards this, the thesis has used the concept of the 'transfer of learning'.

CONTENTS

CHAPTER - 1

GENERAL INTRODUCTION

1.1 Introduction

Examinations are designed and administered at different stages of education. They may be used for classification, grading, guidance, certification and so on. Their particular importance at the tertiary level can not be over-emphasized because, among other things, they define the employment value of the students in the contemporary society. It is this, that has led many to seriously think about the examinations. Various types of reforms over the traditional examination system have been proposed and atleast partly practiced to make the examinations more meaningful and relevant. One of the major steps in this direction is the internal continuous evaluation system.

Whatever may be the flaws in the traditional examination system which is based on a single main examination with an external examiner, it has the advantage of social acceptance. The internal evaluation system with whatever academic advantage it may possess is yet to be fully socially accepted. An examination system in which the teacher on his own gives grades to his students without any external monitoring with respect to the standard of teaching and examination has an important responsibility of continuously establishing its own credibility. The question is not so much regarding the integrity of the teachers, but whether the whole system is academically reliable.

Some of the questions which arise in this connection are whether the examinations are reliable and whether they are valid? Do different examinations in various courses for the same set of students show a reasonable correlation? These questions are important because, in the internal evaluation system, occassionally, a very junior faculty member will be teaching and grading the students in contrast to the external examination system in which usually experienced teachers are chosen as examiners. In such case, then, if examinations are haphazardly planned, their reliability and validity will be limited and they will not be correlated.

Thus, it is important to study examinations and develop methods for their improvements.

It is against the above frame of reference that section 1.2 presents a historical background of the subject matter of examination reforms. This, in section 1.3, is followed by a brief review of the literature on examination reforms. Section 1.4 then presents the objective of the present study, followed by a brief description of the chapter-wise contents of the thesis given in section 1.5.

1.2 Historical Background

Examination System is perhaps older than any other system of education. Teachers have been developing and using their own evaluation procedures for their respective education systems through the ages. Presently, the examinations are becoming extremely vital because people are branded for life on the basis of their examination performance. A clear case in this context is in terms of professional examinations which act as certificates or licençe for people to perform any skill under reference.

As the(examination)system has grown historically, examinations can be broadly classified in terms of categories such as performance-based, oral, written, objective type, short-answer type, long-answer type, traditional essay type, external, internal, continuous, annual, etc.

The system of traditional essay type examinations was first started in China (Harper & Misra, 1976). Those examinations used to last for days and nights. The objective was to identify men for the service of the State. By the 19th century,western countries adopted it for awarding degrees and diplomas. This system got transported to India around the same time when the task was initiated on the Indian soil to re-construct its educational system as per the model of British Universities. In 1854, Wood's despatch recommended the establishment of universities in India. As a result three universities were established at Calcutta, Bombay and Madras. These universities were modelled on the lines of the 19th century British Universities. In this context, there were two options open at that time, viz., the universities of Cambridge and London.Universities in India understandably chose the latter in view of its secular character (Mitra, 1973). University of London offered degree courses in Arts, Law and Medicine. The prerequisite for entering commerce, Arts, etc. was the entrant's success in matriculation examination. The

B.A. degree was for two years and included only one examination. The next higher degree, i.e. M.A., also required two years of study. Thus, the Indian Universities borrowed the London University pattern but with one modification, namely, the introduction of an intermediate examination to be given after two years of matriculation (Singha, 1977). This modification mainly was done with a thought of having a terminal stage of two years after matriculation examination (preparatory stage). As a consequence, owing to the recommendations of the Calcutta University Commission (1917-19), also called Sadler Commission, boards of secondary education started coming up. They had the responsibility of organising curricula and examinations at the secondary stage. The responsibility for all the matriculation examinations were passed on to these boards except in the case of U.P., where the boards were given the responsibility for the intermediate examinations, too. Indeed, this pattern more or less has been in operation to this day.

Briefly speaking, examination reforms have always been uppermost in the minds of educators and, in terms of the Indian educational scene, Ample proof of this is found in reports of various commissions and committees, such as Wood's Despatch (1854), the Indian University Commission (1902), the Calcutta University Commission (1917), the Inter-University Board of India & Ceylon (1924), The Sargent Report (1944), the University Education Commission Report (1950), and finally the Indian Education Commission (1966).

Emphasis on the improvement in examination system became a regular feature only after the report of Indian Education Commission (1881-82). Thus, both Lord Curzon and the Indian University Commission (1902) pointed out the flaws in the system of examinations in the following words:

> "The greatest evil from which the system of university education in India suffers is that teaching is subordinated to examinations and not examinations to teaching".

The Calcutta university Commission (1917-19) also criticized the system of examination, but no positive steps were taken to improve it.

After India became free, the major emphasis on examination reform was first given by the Radhakrishnan Commission (1948) It stressed that "if we are to suggest any single reform in university education, it would be that of examinations". Thus the need for reforming examinations was badly realised and various committees and conferences were organised. As a result, the All India Council of Secondary Education, U.G.C. The Central Ministry of Education, the Central Evaluation Units and several state education departments discussed from time to time the issue of examination reform, realising that the present system of examination was a great obstacle in the educational process.

In 1965, the U.G.C. appointed a committee to look into the problems of university examinations. This was followed by the Kothari Commission (1964-66). This commission, like the previous ones also felt the need for reforming examinations. One of the important documents regarding examinations came from the U.G.C. in 1975. It was titled "A Plan of Action" and it attempted to provide a practical way for reforming examinations in the Indian universities.

The above then is an overview of the theme of 'examination reform' in its historical context. As can be seen from the details given, this subject matter has always been of interest to the various sections of the society. And, it is in terms of this universality of its that this area has been receiving a detailed attention of the researchers in the field of education.

1.3 A Review of the Literature on Examination Reforms

A review of the work done on reliability and validity of examinations will be discussed in Chapters 3 and 4, respectively. Thus, this section will mainly concern itself with the work in the area of internal assessment.

Traditional essay type external examinations have been criticised on many occasions. As a result, internal assessment system has been implemented in a number of universities from abroad as well as from India; providing a ready canvas for the educational researchers to explore the efficiency of this alternative system of evaluation.

Indeed the views and opinions of educationists and researchers on the subject of internal assessment cover extremes of every type. Thus, while, Secondary Education Commission (1952), the Ministry of Education (1956), Gayen et.al.(1961), the Secondary Education Committee, Assam (1965), Rao(1970), Singh (1970) and Gayen (1970) and several others have recommended that in the traditional system of external examination some weightage should be given to internal assessment type evaluation, Raina (1967) has expressed a contrary view in this matter.

In the above context, several other researchers have studied the subject of internal assessment at the macro-level as also at the micro-level. Thus, Gayen et.al.(1962) and Lele et.al. (1962) observed variation in external and internal assessments, and suggested that marks of both the examinations should be scaled before they are added up. Taylor (1962) holds that the two assessments should not be combined unless they are checked for their accuracy. Gayen et.al. (1962) observed that the total of internal plus external assessment is more reliable than taking them separately. Misra (1973) stated that the internal assessment is more consistent from year to year than the external assessment. Further, Natarajan (1977) has argued that combining the two sets of scores (namely, those obtained in internal and external examinations) is not an academically sound procedure, as these two forms of assessment test two different sets of abilities and thus suggests that it is therefore important to assess them separately and also show them separately.

Singh (1971) in the article titled 'Examination Reform: Internal Assessment' discusses the concept of internal assessment, its advantages and the problems related with it. Lingamurthy (1970) states that, along with memory, proper evaluation of students' personality traits should also be done. He observes that 'along with Jøyanam, Vigyanam should also be tested'. Kahlon (1977) in his paper 'Student Evaluation and Internal Assessment' talks about the following objectives which can be achieved through an internal system of evaluation in the manner given below:

1. It assists in the evaluation of personality characteristics, in addition to those which are needed for the success in a chosen field.

2. It gives an accurate comparison of an individual performance vis-a-vis that of the others.

3. It improves the basis of prediction of success in the educational, occupational and professional spheres.

4. It identifies the student's capacity, his potentiality as well as his limitations.

Kahlon further states: "Techniques employed for students' evaluation should achieve these objectives and build up the confidence in the mind of a student about the objectivity of the system".

SCERT, Tamilnadu (1976) in collaboration with NCERT presents a 'Scheme of Comprehensive Internal Assessment and Manual of Instructions'. This scheme attempts to bring into focus the important non-scholastic achievements in personal and social qualities like regularity, punctuality, initiative, cooperative spirit, sense of responsibility, industry, civic consciousness, discipline and a spirit of social service in addition to interests and attitutdes in literary and scientific spheres. It also takes into consideration participation and performance in the co-curricular field such as scouting, NCC and cultural activities.

The scheme observes: "Any evaluation, if it is to be scientific should be developed as the right tool to assess the versatile capabilities of the pupils. It should never be designed as a trap or a man-hole to corner and catch the pupils at their weak spots". The aim of any evaluational programme should be to identify and assess the original potentialities in the pupils.

One of the important document regarding internal assessment is brought out by the Association of Indian Universities called, 'Monograph on Internal Assessment for Universities'. The basic question posed here is 'who should make a design for the internal assessment'. The answer given is that 'the teacher, who teaches the student, would be best for this purpose'. This monograph further investigates the design aspect of internal assessment. For this purpose a questionnaire was prepared and the consensus was worked out based on the opinions of the teachers. The monograph includes this aspect in the section on 'Design of Internal Assessment'. In addition to this, the monograph includes the theory of continuous internal evaluation, the need for it in the Indian context, its basic principles, the mechanics of continuous internal assessment and, finally, the issues of proper recording and storing machinery for the successful implementation of this evaluation scheme. It also contains a good bibliography on the subject of internal assessment.

The above then is a brief review of the literature on the subject of 'internal assessment' as pursued particularly in the Indian context. As can be seen, much of the work is particularly in terms of partly-external-partly-internal type evaluation situations which seem to have received much more attention of the researchers than the examinations coming under what is often termed as the 'totally internal system of evaluation'.

1.4 Objectives of the Present Study

This thesis is devoted to the study of internal continuous evaluation system at the tertiary level. One of the main investigations aimed at through this thesis pertains to the reliability studies for the examinations of quizzes, tests and comprehensives, as conducted under the internal system of continuous evaluation. Along with this, as a natural corollary, the thesis also concerns itself with the theme of 'course-reliability' as also with the correlation studies between various examinations or evaluation components of quizzes, viva, seminar, tests, project-work, group-discussion, comprehensives, etc.

Another main investigation aimed at through this thesis concerns the validity studies for the examinations under the internal system of continuous evaluation. Here the objective is to delineate a technique for such an investigation through a detailed illustration. Thus, in this context, the thesis studies the theme of predictive validity, taking the student performances in the courses of 'Concepts in Science' and 'Modern Physics' as predictors, the other host of student performance indices as available in an internal system of continuous evaluation being candidates for the role of criterion. A similar study is also undertaken with the student performance in Higher Secondary playing the role of a predictor; thus, in turn, making it possible for the investigation to compare the student performances in Higher Secondary, Concepts in Science and Modern Physics for the choice of the best predictor.

Another important aspect of the validity study aimed at in the above context concerns the theme of the 'concurrent validity'.

In addition to the above stated reliability and validity studies, the thesis also aims at investigating teachers' perception of the objectives of education under the internal system of continuous evaluation at the tertiary level. Further, the thesis also studies teachers' perception and, wherever feasible, practices as followed by them in terms of objectives, techniques and nature of examinations in the above context.

Further, as another of its main investigation, the thesis also aims at studying the internal continuous evaluation system as under the 'Practice School (PS) method of education'. As an overview, the PS can best be described as an attempt to build the much needed bridge between the professional world and the academic world. Thus, just as a medico undergoes an internship before graduation, similarly the PS system of education requires students of engineering, science and humanities to practice their profession during their educational years. The most important feature of the PS system of evaluation is in terms of the fact that, consistent with the theme of internal assessment, the entire student education and evaluation under PS is the total responsibility of the university faculty resident at the university's PS centre. In the context of this investigation, the thesis first concerns itself with presenting the themes of 'consumer obligation' and 'multiple objective' testing, which are central to the PS system of education, and then devotes itself to study some of the reliability and correlation aspects for examinations under PS. Another important investigation that the thesis aims at in this context pertains to the question of the content validation of the PS evaluation.

Finally, in the context of the educational innovation of the Master of Engineering (M.E.) (Collaborative) programmes, which constitute a natural corollary to the PS system of education, the thesis applies itself to the investigation as to what can be the model of a internal continuous evaluation system, if the multiple objective evaluation were to be incorporated in the class-room based teaching effort.

The above then is a comprehensive statement of the objectives as pursued under the present study.

At this stage, it may be mentioned that, towards the information and the data base required for the above stated objectives, this thesis heavily draws from the educational operations at the Birla Institute of Technology & Science, (BITS), Pilani. As a 'deemed university', committed to the theme of 'educational innovation', right since its inception, BITS has adapted the semester pattern of total internal continuous evaluation, characterized by letter grading. Its educational system, aiming at interdisciplinary student training, is broad-based but integrated in character, with heavy emphasis on the analytical techniques. Since the academic year 1972-73, BITS has also introduced across its disciplines of engineering, science and humanities the Practice School System of edmation. And, since the academic year 1979-80, BITS has initiated the M.E.(Collaborative) programmes at its off-campus educational centres. Further, as an all India Institute, BITS attracts students from all parts of the country. Similarly, its faculty is also of all India character, and reflects all shades of higher education, from India as also from abroad. It is in this context that the thesis has then looked upon the educational operations at BITS as on-going educational experiments for generation of the much needed data-base for the pursuing of the purposed educational research.

1.5 Summary of Chapters

It is against the frame of reference as described in the previous section that Chapter 2 basically deals with finding teachers' perception towards objectives of education and examinations in continuous internal evaluation system. Various other aspects of the continuous internal evaluation system are discussed in chapters which follow.

Thus, Chapter 3 gives a study of reliability of various classroom examinations and courses and also correlations between various examinations.

Further, Chapter 4 concerns itself with the criterion related validity of certain class-room examinations.

Chapter 5 provides a brief description of the Practice School (PS) system of education and, in this context, introduces the concept of 'consumer obligation' and 'multiple objective assessment'.

It is in the above context that Chapter 6 then proceeds to study some of the reliability and validity aspects for PS evaluation in terms of reliability of two evaluation components, namely, Viva and Quiz; examiner reliability of several components course reliability; inter-correlations between various components of PS evaluation; and, finally its content validity.

Chapter 7, using the data base as available from the recently introduced ME(Collaborative) programmes at BITS, concern itself with theme of evolving of a model for a multiple-objective assessment with reference to the class-room based teaching effort

Finally, Chapter 8 concludes the thesis. This is then followed by various Appendices.

Needless to say the details of the above chapters have from time to time involved data processing activity. Towards this the thesis has extensively used the IBM 1130 digital computer. The programmes have been pursued in the Fortran language. These have been given in Appendix I .

CHAPTER - 2

ON THE OBJECTIVES OF EDUCATION AND EXAMINATIONS

2.1 Introduction

It is important to define the objectives of education before an efficient system of education can be designed. To achieve this goal, the educator, according to Tyler (1949), will have "to first determine the objectives which he seeks to attain, select the most appropriate learning strategies to obtain these objectives and organise their sequencing, and, finally, evaluate the extent to which they have been obtained". Thus, closely linked with the objectives of education is the objective of examination. If the examinations are to play the desired role as an integral part of the educational system, it is of paramount importance that the objectives of the examinations should be precisely defined and linked with the objectives of education. This linkage or integration of examinations with the educational system should be relatively easy to achieve in an internal evaluation system where the teachers are also examiners. This chapter makes an attempt to find out directly from the teachers their own perception of the objectives of education and the objectives of examination in an internal evaluation system. For this purpose, teachers involved in such a system as at BITS have been selected as a data-base. Further, this chapter also states the various techniques of evaluation, i.e., quizzes, tests, home assignments, etc., used by the teachers to achieve the objectives of evaluation.

The chapter has eight sections. Section 2.2 discusses the need for defining objectives of education. The formulation of objectives is given in section 2.3. Some work on determination of the objectives of education is given in section 2.4, whereas section 2.5 provides the objective of the present study. The chosen sample and method for data collection have been described in section 2.6 and the results with their analysis are presented in section 2.7. Finally, the conclusions of the study have been summarised in section 2.8.

2.2 Need for Defining Objectives of Education

The basic tasks of an educator listed by Tyler (1949) are as follows:

(1) Determination of the objectives which the course, curriculum (or the institution) should seek to attain.

(2) Determination of the extent to which the objectives are attained.

The first task listed by Tyler relates to the statement of objectives of the curriculum and course-work. The second task, of course, is linked with the statement of objectives of the evaluation/examination process.

It is therefore important to evaluate the extent to which objectives of the curriculum have been achieved and this, as is obvious, can only be done by a proper assessment/examination. Heywood (1977) while discussing the educational system observes: "Despite the fact that it is generally agreed that assessment procedures have a powerful influence over learning, little has been done to integrate them into the curriculum process. There is a real danger that techniques and processes (e.g.

continuous assessment) will be introduced without reference to the objectives they are supposed to obtain. As a consequence, they may be unsuitable for evaluation/(examination) because they may not be valid measures of the objectives which it is hoped will be tested. We cannot expect to improve the educational process without adequate evaluation/(examination) of learning. We cannot have effective evaluation/(examination) without some form of assessment/(examination) carefully related to specified objectives and learning strategies designed to obtain those objectives".

Thus, with suitably defined objectives, it would become easier for educators to plan the course work, choose suitable learning techniques and, finally, plan proper examinations to evaluate whether the desired objectives of education have been achieved.

2.3 The Formulation of Objectives

The next significant question which arises is how the objectives should be formulated. This has been given in detail by Bloom (1979a) in his Taxonomy of educational Objectives, Volume 1. Defining educational objectives means formulating the ways in which students are expected to be changed by the education process. The changes may be in their thinking, their feelings and their actions. "There are many possible changes which can take place as a result of learning process but since time and resources of the institutions are limited only a few possibilities can be realised", (Bloom, 1979a). Bloom further remarks that, it is important, therefore, "that the major objectives of the school or unit of instruction be clearly identified if time and effort are not to be wasted on less important things".

In an internal evaluation system the teacher has to play different roles. He is the curriculum designer, teacher and evaluator. Hence, in such a system the teacher is an important link in the chain of formulating objectives. The integration of examinations with the educational system should be relatively easy to achieve in an internal evaluation system where the teachers are also examiners.

Some of the major criticisms of examinations pertaining to ill defined or inconsistently defined objectives are described below:

1. Examinations are not designed around clearly defined objectives. There is not a close relationship between curriculum, learning experience and the assessment/examination procedure. They are seldom used as a part of an integrated educational process. Indeed, one can never have effective evaluation unless examinations are related to specified objectives and learning strategies designed to obtain those objectives. Heywood (1977) has emphasized this by stating that, "A positive approach to examinations is likely to have powerful and beneficial effect on learning. To achieve this goal, evaluation must become an integral part of the teaching activity.

2. The evaluation (examination) should begin by obtaining a clear understanding of what it is that one wishes to achieve. Some of the variance in the student assessment is undoubtedly due to the fact that there are unconscious differences of

opinion amongst examiners about the objectives of the examination. It has been reported by many studies (Bloom, 1979a) that quite often the examiners are not consistent about the objectives of examination, the fulfilment of which they expect out of their students.

3. Examinations do not assess the objectives they intend to measure. Indeed this is the aspect pertaining to the validity of the examinations.

2.4 The Literature Review

The literature on higher education contains quite a number of studies on the identification of the objectives of education. As stated by Heywood (1977) the most common method of obtaining the objectives is to make a 'survey'. Though other methods e.g. 'Delphi' and 'Task analysis' given by Heywood can also be adopted, the 'survey' seems to be the most convenient and is hence common. Detailed listing of the objectives of education along with a comprehensive description of ~~the~~ each objective is given in 'Taxonomy of Educational Objectives' by Bloom (1979a).

It would be useful to discuss the review of the work done by various researchers on objectives of education under the following heads:

- Researches reported on finding of educational objectives.
- The procedure for obtaining objectives: The Survey Method
- Review of the comprehensive study on objectives: Taxonomy of educational objectives.

2.4.1 Researches which reported on finding of educational objectives

A comprehensive review of the work on the identification of objectives of education has been made by Heywood (1977).

Beard, Healy and Holloway (1974) observe that quite a number of studies for various subjects/disciplines have been carried out in this area, but the overall impact of the thinking by the educational researchers is small and seems to have been generated mainly in the science and technology part of the spectrum. According to them, studies in the other disciplines are few compared to those in science and technology. As regards to such studies for the medical education, the same have been carried out by Miller (1962), McGuire(1967),Russel and Walton (1970), Freeman and Byrne (1973) and W.H.O.(1974). According to them, "Medical profession has concerned itself with objectives in the strictly behavioural sense and seems to have been the only profession to have really set about this activity in the formal style of curriculum evaluation".

The literature reports an interesting effort in the direction of evolving the educational objectives. In specific terms, a course in educational technology described by Erikson (1968) gives students an introduction which relates education to the structure of society and teaches methods to determine educational objectives. This is what a sound educational system should do .

Aims and objectives of education have also been worked out in a number of reports by UNESCO (1966), Mac Ewan(1970), Long, Meltzer and Hilton (1970), Sheffield (1970), Cowan(1971), Perren and Trim (1971), Hull (1971), Wain (1972), Vaugham and Rouve (1972), and Carter and Lee (1974).

As stated earlier, there have been more studies on finding objectives of education in the field of science and technology as compared to the humanities. However, as quoted by Heywood (1977), some important work in the area of humanities has been carried out by researchers like Yudkin (1954), Plumb (1964), and Kaysen (1973).

Indeed, the formulation of objectives in education is not the end of the problem but the selection of the most important objectives is also a major issue to be considered. Dressel (1965) points out this fact and states that quite often teachers formulate more objectives than can be achieved by the students. He suggests that it is better to select a few well chosen objectives. Bloom (1979a) also has a similar view point. To overcome the problems regarding formulation of objectives, priorities of the objectives to be incorporated in education will have to be set up. This is necessary because the time and resources available to the institutions are limited.

Priorities of aims and objectives, as Heywood (1977) believes, may be accomplished by rating objectives against the philosophical and social aims of the curriculum. Which objectives are likely to achieve these aims? Which of these aims are important? These may be some of the questions which couldbe of help in deciding the priorities among aims and objectives.

2.4.2 <u>The Procedure for obtaining the Objectives: The Survey Method</u>

Haywood (1977) writes, "The survey is probably the most popular method of obtaining information, particularly about the needs of professional people in relation to the syllabus." Further, he says, "Because of its simplicity, the survey is an extremely useful way of obtaining fact and opinion". In recent years several researches have obtained through the survey technique ratings of objectives from professional people, teachers and students. Indeed, Heywood (1977) points out a danger in such an approach in the sense that, "opinions will be (may get) accepted as a fact without further qualifying tests". But then, leaving aside such drawbacks, which can be taken care of through proper further examinations of the data-base collected, the survey has been the most popular tool for obtaining information. The study now quotes some of the important educational researches based on surveys which may be of use as far as the subject matter of finding objectives is concerned.

Lee (1969) asked a sample of mechanical engineers to rate twenty behavioural objectives of laboratory work. Also he asked teachers to indicate by rating which of the best two teaching methods would be most suited in order to achieve these objectives. They were also asked to give their opinions about the four most important objectives. Results of the survey conducted by Lee emphasize the significance of affective domain to students at tertiary level. Teachers also suggested the dominance of objectives from 'affective domain' over 'cognitive domain'. The terms 'affective' and 'cognitive' would be discussed in the next sub-section.

Burke (1969) has reported that in the University of Lancaster a group of staff and students used to meet regularly between 1965 and 1967 to discuss problems arising in the university. Such discussions were helpful in deciding the aims and priorities of education.

Dew-Hughes and their co-workers (1966) had similar types of encounters with the students. They surveyed opinions of all the members across the university. Each and every person of the university, whether teacher or student, was asked to list the qualities he would like the examination system to assess. Many suggestions overlapped and finally the qualities that emerged with maximum concensus are given below:

All of them agreed to the suggestion that the examination system should evaluate 'academic ability'. What they meant by 'academic ability' has been listed as.

Creativity: absorption, assimilation, synthesis and application.

Objectivity: transferability of values and methods.

Perception: Analysis, speed and depth.

Knowledge: Knowledge of facts or of sources, memory and understanding.

Coherence: Communication, both written and spoken.

Rest of the suggestions were on personality factors which affect academic ability. These factors are:

- motivation and interest in the subject
- Outside interests
- diligence
- ability to work on your own and with others.
- responsibility to one's own ideas, and to oneself and to others
- behaviour under stress - clarity of thought and reaction under challenge.

2.4.3 Review of the Comprehensive Study on Objectives Taxonomy of the educational objectives

A comprehensive work on objectives of education is found in the "Taxonomy of "Educational Objectives" by Bloom (1979a, 1979b) This section gives a very short summary of this work.

Along with the development, mainly at the school level, of objective testing, an attempt to design at the tertiary level test items to test high level skills such as ability to plan, analyse, judge, etc. has come up. These skills are called behavioural objectives and along with the various objectives of the tests/examinations at the various levels of schooling go to form the universe of educational objectives. When the outcomes of learning are evaluated in terms of its objectives, it is termed as an "objective approach". The taxonomy of educational objectives by Bloom is a major breakthrough in the field of application of the objective approach to the problem of studying the reliability and validity of the examinations at the tertiary level. Though the work of Bloom centres around the details pertaining to a specific group, due to its comprehensiveness, it has become quite universal, as it covers the entire spectrum of educational objectives cutting across all the levels of education.

The taxonomy is published in two volumes: Volume 1 deals with the 'cognitive domain' and Volume 2 with the 'affective domain'. Bloom defines these broader areas as follows:

(I) Cognitive Domain: It includes, "those objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills".

(II) Affective Domain:

It, "includes objectives which describe changes in interest, attitudes and values, and the development of appreciations and adequate adjustment".

This taxonomy has been constructed by a group of educationists and is an attempt to define the skills of learning. They believe that skills are hierarchically ordered. In cognitive domain the skill of 'ability to evaluate alternatives' is at the top of the hierarchy, while 'knowledge' is at the bottom. The main titles of the skills have been listed in the following table:

TABLE 2.1

Titles of Main Skills in the Cognitive and Affective Domains indicating their Order of Hierarchy

Main Skills from the Cognitive Domain in the Increasing Order of Hierarchy	Main Skills from the Affective Domain in the Increasing Order of Hierarchy.
1. Knowledge	1. Receiving
2. Comprehension	2. Responding
3. Application	3. Valuing
4. Analysis	4. Organisation
5. Synthesis	5. Characterisation by a value or value complex.
6. Evaluation	

The working group of the "taxonomy" has defined a third domain other than 'cognitive' and 'affective' and has named it as 'psychomotor domain', which, according to Bloom (1979a), "is the manipulative or motor-skill area". Apparently, this field of study, as of today, is yet to receive any detailed attention of the educational researchers.

2.5 Objective of the Present Study

The above then is the state of the art in terms of the research in the area of determining of the educational objectives. Indeed, as can be seen from the details given, not much work is reported in India in this area and certainly there is no study in terms of such investigation for the internal system of continuous evaluation at the tertiary level. It is against this background that, in the pages to follow, this chapter investigates, as a case study, the education system at BITS, with a view to understand the aims and objectives of an education system characterised by the internal system of continuous evaluation under the semester pattern of the course offerings.

The present study has been done in two parts; the objectives of which have been stated separately.

2.5.1 Part One: "Determining Objectives of Education"

The objective of this part of study is the find out the opinion of teachers about the qualities that a successful person should have. "Should these qualities be imparted through education, or, can these qualities be passed on through an educational system like that of BITS" are some of the questions on which the opinions of teachers were sought. The last question pertaining to study tries to find out how far the teachers have been able to incorporate in education the qualities which are also very essential for success in life.

2.5.2 Part Two: "Determining Objectives of Examinations"

This part of the study aims to establish the objectives which examinations in an internal continuous evaluation system are endeavouring to accomplish. It also aims to find out the techniques of examinations such as, quizzes, tests, etc., that are included in evaluation in order to attain the objectives. The final section of this part enquires about the types of question-papers (objective, short-answer, etc.) that are designed in order to achieve the objectives at hand. Thus, this part of the study is solely concerned with the objectives of the examinations.

2.6 Sample and Data Collection Procedure

Towards the proposed study, a survey was conducted by giving two prepared questionnaires to the teachers involved in the internal continuous evaluation system. As mentioned earlier, the teachers of BITS were selected as the data-base because the faculty are from all over India, trained abroad as well as in India, and belong to all the disciplines and are aquainted with all the levels as also the patterns of education. Hence, the faculty at BITS provided a suitable sample for the study.

First questionnaire which deals with the determination of objectives of education is given in Appendix A, whereas the second questionnaire dealing with the determination of objectives as also the nature and techniques of examinations is given in Appendix B.

The procedures for conducting the survey for the two above mentioned questionnaires along with the description of the corresponding faculty compositions which responded to it (i.e. survey) is given below.

2.6.1 The sample and Data Collection Procedures for the Questionnaire I on 'Determining the Objectives of Education'

For the purpose of the interview the researcher personally contacted the teachers. The total sample thus emerging consisted of 51 teachers drawn from across the Institute's disciplines. The sample is described in Table 2.2.

Table 2.2

Teachers from Different Disciplines who Responded to the Questionnaire I Supported by the Interview

Discipline of the Teacher	Number of Teachers
Science	26
Engineering	12
Humanities	10
Management	3
Total	51

Thus, it can be seen from table 2.2 that almost 50% of the teachers who responded came from Science disciplines, while the remaining 50% came from the disciplines of Engineering, Humanities and Management.

The purpose of this survey study being to find out the objectives of education, the questionnnaire I had enumerated some of the qualities which are common to most adult situations, considered essential for living. The teachers were requested to rate/rank these qualities. Further, the teachers were also asked questions in term of which of these qualities the process of formal education should try to achieve, which can be achieved and, finally, which have been achieved. The answers to these questions were sought under 'yes' or 'no' type categories.

2.6.2 The Sample and Data Collection Procedures for the Questionnaire II on 'Determining the Objectives, Nature and Techniques of Examinations'

Like in the case of the Questionnaire I, the Questionnaire II was also distributed to all the teachers across the Institute. The total number responded to was 72. While answering the questions of the questionnaire II, each teacher had kept in mind the examination details pertaining to a given definite course of his choice that he has been teaching over the years. Table 2.3 gives the teacher's-own-discipline-wise as also the course-discipline-wise categorization of the responses received.

Table 2.3

Teacher-Discipline-wise as also Course-Discipline-wise statistics for the Responses Received for the Questionnaire II

Discipline of the course \ Teacher's own Discipline	Science	Engineering	Humanities	Management
Science	29	1		
Engineering	4	19		2
Humanities	1		13	
Management		2		1
Total	34	22	13	3

The questionnaire II was so designed as to be self-explanatory. However, whenever needed, the researcher also personally contacted the faculty to elaborate on the information sought through the questions posed.

The questionnaire in all had three parts, namely, part A, Part B and part C. These parts are briefly described below:

Part A: On Determining Objectives of Examinations

This section of the questionnaire included a list of objectives of evaluation based on Bloom's Taxonomy of Educational Objectives as coming under the cognitive domain. The objectives considered were knowledge, comprehension, application, analysis, synthesis and evaluation. The researcher with the help of the supporting illustrations explained to the teachers in detail as to what was meant by these objectives. Clarifications regarding the objectives were also given in the explanatory notes which were included with the questionnaire. One of the questions posed to the teachers under this part of the questionnaire was which of the objectives did they incorporate in their evaluations. Teachers were to answer this question separately for quizzes, tests and comprehensives. Answers were sought in terms of an indication as to what percentages of the total assessment as aimed at through the examination corresponded to the different objectives stated to be pursued under that examination. Needless to say, to answer this question, then, the teachers had to recall the question papers for the courses against the examination details of which they were answering the Questionnaire II. As a result, it may be mentioned that almost all (i.e. about 95%) of the teachers had kept the respective question papers before themselves while answering the questionnaire. Thus, in the context of the examination paper under consideration, a teacher could arrive at the various percentages for the different educational objectives corresponding to the said examination. by studying in detail the educational objectivess for each of the test items pertaining to the questionpaper.

Part B : On Determining Techniques of Examinations

This section of the questionnaire was related to the techniques of evaluation as practiced by the teachers. Here the questions posed to the teachers was, 'How many quizzes and tests do you conduct in the course in a semester'? The teachers were asked to indicate the number of quizzes and tests and also of 'any other' components of evaluation along with the weightages assigned to them, constituting the total evaluation scheme for the course under consideration. Other components of evaluation could be home assignments, lab work, project work, viva, seminar, group discussion and so on. Further, information is sought on, what objectives teachers are trying to achieve through these other components of evaluation

Part C : On Determining Nature of Examinations

Part C of the questionnaire contained questions about the nature of examination system, viz., whether the examination was closed-book, open book or a combination of both. Another question asked was "whether the examinations permitted any choice of questions/marks or not". The teachers' opinion on giving of the choice of questions in the examinations was also sought. The last question in this part was about the type of questions included in the examinations. Based on the different types of questions, the examinations were broadly classified in terms of following categories:

1. Objective Type

The objective type test includes purely multiple choice questions. For example, the questions can either be true-false or choose-the-correct-alternative type or fill-in-the-blanks type, and so on.

2. Short-answer Descriptive

As the name indicates, questions in this case require short answers and are memory-oriented.

3. Short-answer Problem Solving

This category requires short answers and the questions are application-oriented.

4. Long-answer Descriptive

Questions are memory-oriented and require lengthy answers.

5. Long-answer Problem Solving

Answers are lengthy and application-oriented.

2.7 Results and Analysis

2.7.1 Results and Analysis Towards Determining the Objectives of Education

As described in the previous section, the questionnaire I was concerned with the objectives of education. There were four questions in this questionnaire which centered around the

TABLE 2.4

Teacher Responses Indicating Ranks for Various Qualities listed in Question I of the Questionnaire I As to Their Perception of the Importance of the Qualities in life.

Rank	Perception	Analysis	Diagnosis	Problem Solvings	Judgement	Communication	Understanding	Sympathy	Tolerance	Sense of Responsibility	Leadership	Decision Making	Coping with Frustration	Knowledge	Creativeness	Self Reliance	Confidence	Sense of Humor	Ability to Mix Well	Wide interest	Independence	Honesty	Ambition	Common Sense	Logical Thought
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	5	3	7	1	2	2	5	9	4	2	8	1	10	7	4	3	2	5	4	3	4	7	7	2	6
1.0	2	1	0	2	1	0	1	0	1	7	0	2	1	4	0	6	4	0	0	1	1	9	0	1	4
1.5	0	0	0	0	0	0	1	0	0	3	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0
2.0	5	0	3	2	2	1	3	4	3	3	0	3	1	1	2	3	4	1	0	0	0	2	2	2	0
2.5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0
3.0	1	4	1	1	1	2	4	2	3	3	1	2	0	3	0	1	5	0	1	1	1	0	0	1	2
3.5	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1
4.0	1	4	0	3	2	0	5	2	0	3	1	2	0	0	0	4	2	2	0	0	0	0	1	2	1
4.5	1	0	0	0	0	0	1	2	1	1	0	1	0	1	1	0	1	0	0	1	0	1	0	2	0
5.0	4	1	3	1	1	3	0	1	1	0	3	3	0	2	4	0	2	0	2	2	2	1	0	1	1
5.5	1	1	1	1	0	0	1	0	1	1	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0
6.0	4	5	2	5	6	2	3	0	2	1	0	5	0	1	2	2	1	0	3	1	2	0	2	5	5
6.5	2	1	0	1	2	2	2	1	0	2	0	2	0	0	2	0	1	0	0	0	0	3	1	2	2
7.0	2	3	5	3	5	3	4	0	1	2	2	5	2	4	0	2	3	0	2	1	3	3	2	6	2
7.5	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8.0	2	4	1	1	3	3	0	1	0	1	1	4	1	1	4	3	3	3	0	0	2	1	1	2	2
8.5	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	1
9.0	3	0	3	0	2	1	0	0	2	2	2	3	1	3	3	0	1	1	1	1	1	0	0	3	3
9.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
10.0	0	1	1	1	2	2	1	2	0	4	1	1	1	1	3	3	2	1	1	0	0	1	3	0	2
10.5	1	0	0	0	1	1	1	0	0	2	1	0	1	0	0	2	1	0	0	0	1	0	1	0	1
11.0	2	0	1	2	2	1	1	0	1	3	2	1	3	2	4	3	0	2	2	1	2	2	0	1	4
11.5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

0 Rank refers to no rank assigned.

......2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
12.0	1	3	0	1	0	1	0	0	2	2	5	0	0	0	1	3	3	2	0	4	1	1	1	2	1
12.5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0
13.0	3	0	3	0	2	0	1	1	2	0	0	2	1	2	1	1	2	0	2	1	1	3	4	2	0
13.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.0	0	2	3	2	2	2	2	2	1	1	3	0	0	1	1	0	1	2	0	3	2	2	3	1	1
14.5	0	3	1	1	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	1	0	0
15.0	0	2	1	1	2	3	2	0	3	1	1	1	3	1	2	1	4	1	1	2	0	2	3	2	2
15.5	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0
16.0	2	1	2	6	1	2	2	1	2	1	2	3	1	0	1	3	3	1	2	2	4	0	1	1	1
16.5	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0
17.0	1	2	1	2	2	4	2	3	1	2	2	2	1	5	3	2	0	4	2	3	5	2	1	4	4
17.5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
18.0	1	1	0	1	2	3	1	3	2	0	2	1	0	0	1	1	0	1	4	2	1	0	4	2	1
18.5	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
19.0	1	1	2	1	2	3	3	1	4	1	0	0	1	0	1	0	2	2	4	3	1	0	0	1	2
19.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
20.0	0	2	1	2	0	0	1	5	3	1	0	1	1	0	0	2	0	3	1	1	1	1	2	2	0
20.5	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	1	1	1	0	1	0	0
21.0	0	1	1	0	1	4	0	1	2	1	2	0	3	4	3	0	0	2	3	2	1	2	0	0	1
21.5	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22.0	1	1	0	0	0	2	0	2	1	0	2	2	3	3	3	1	1	3	3	2	3	1	4	0	0
22.5	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
23.0	0	1	1	1	1	3	0	0	0	0	2	0	4	1	0	0	0	3	4	3	3	1	2	2	1
23.5	0	0	0	1	0	0	0	2	1	0	1	0	.1	0	0	0	0	0	0	0	0	0	0	0	0
24.0	1	1	2	1	1	0	0	2	0	0	4	0	0	0	1	0	0	5	2	6	2	0	0	0	0
24.5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
25.0	1	0	1	2	0	0	0	1	2	0	0	0	8	2	0	0	0	2	1	1	3	2	1	1	0

objectives of education. The results and analysis for these four questions are separately discussed below:

(I) Question I: Results and Analysis for the Qualities Considered Important in Life.

As mentioned earlier, the teachers were requested to rank the already listed set of qualities as per their (teachers') visualisation of the importance of those qualities in life. The results, i.e., the teacher responses obtained in this context are given in Table 2.4.

For the purpose of analysis, the teacher responses are processed as follows:

(i) Defining of the rank-wise Categories and Assigning of Weights to them

All the ranks have been classified into six categories. Further, a weightage has been assigned to each category as shown below in table 2.5.

Table 2.5

Description of the Weightages Assigned to different rank-wise categorizations

Rank-wise Categorization	Weightage Assigned
Categorization C_1 for ranks between 1 to 5.5.	$W_1 = 10$
Categorization C_2 for ranks between 6 to 10.5	$W_2 = 8$
Categorization C_3 for ranks between 11 to 15.5	$W_3 = 6$
Categorization C_4 for ranks between 16 to 20.5	$W_4 = 4$
Categorization C_5 for ranks between 21 to 25	$W_5 = 2$
Categorization C_o indicating 'no rank' assigned.	$W_o = 0$

Further, as can be seen from table 2.4, some of the qualities occupy partial ranks say, 5.5. Procedure for arriving at such a rank (5.5) is as follows:

For example, a teacher gave rank 5 to two qualities say, perception and analysis. For the purpose of analysis, instead of ranking one quality as 5 and the other 6, or both 5 or both 6, the researcher simply averaged the ranks and gave each above mentioned qualities a rank of 5.5.

(ii) <u>Calculation of the 'Average-Importance-Index'</u>

To obtain the 'average-importance-index' for a given quality as emerging from the teacher responses, first a frequency distribution corresponding to the quality under consideration is obtained indicating how many times the 'quality' has responses under the given rankwise category as listed above. Thus, for, say, the i^{th} quality one gets six frequencies (corresponding to the six above listed rank-wise categories from C_1 to C_5 and C_0), namely, f_{i1}, f_{i2}, f_{i3}, f_{i4}, f_{i5} and f_{i0}.

Then, the average importance-index, $\bar{I}_i$, for the i^{th} quality is given by:

$$\bar{I}_i = \frac{\sum_{j=0}^{5} f_{ij} w_j}{\sum_{j=0}^{5} f_{ij}} \qquad 2.1$$

Where, f_{ij} = frequency of responses for i^{th} quality in terms of it being assigned j^{th} rank-wise category, and

w_j = weightage, assigned to the j^{th} rank-wise category.

(iii) Obtaining the mean, $\bar{I}_Q$, as also the standard deviation, σ_Q, for the sequence of the average-importance-index $\{\bar{I}_1\}$

Next, for the sequence of the average importance index $\{\bar{I}_1\}$, ~~are calculated~~ the arithmetic mean, $\bar{I}_Q$, and the standard deviation σ_Q are calculated.

(iv) Categorization of Qualities based on $\bar{I}_Q$ and σ_Q:

It is on the basis of the values of $\bar{I}_Q$ and σ_Q that then the various qualities as per their importance or significan[ce] or priority level are classified as follows:

Table 2.6

Quality Priority Level Description

Range of the Average Importance Index	Description of the Quality importance/significance/priority level.
$(\bar{I}_Q+3\sigma_Q)$ to $(\bar{I}_Q+2\sigma_Q)$	Highest priority
$(\bar{I}_Q+2\sigma_Q)$ to $(\bar{I}_Q+\sigma_Q)$	Good priority
$(\bar{I}_Q+\sigma_Q)$ to $(\bar{I}_Q)$	Average priority
$(\bar{I}_Q)$ to $(\bar{I}_Q-\sigma_Q)$	Less than Average priority
$(\bar{I}_Q-\sigma_Q)$ to $(\bar{I}_Q-2\sigma_Q)$	Least priority

The results of table 2.6 are then analysed along the above stated approach and the emerging priority-level wise categorization of the qualities considered important in life is reported in the 1st column of the table 2.9.

(II) Question 2: Results and analysis of Qualities vis-a-vis the Question as to which of these qualities the education should try to achieve

As indicated in section 2.6.1, the responses requested

TABLE 2.7

Teachers' Responses in terms of Percentages of 'Yes' for Questions 2, 3 and 4 of Questionnaire I

Questions posed before the teachers / Qualities	Qualities that should be achieved	Qualities that can be achieved	Qualitie that you have been able to achieve
1. Perception	78.43	72.55	62.75
2. Analysis	72.55	62.75	86.27
3. Diagnosis	62.75	86.27	90.20
4. Problem solving	86.27	90.20	76.47
5. Judgement	90.20	76.47	76.47
6. Communication	76.47	76.47	76.47
7. Understanding	76.47	76.47	52.94
8. Sympathy	76.47	52.94	94.12
9. Tolerance	52.94	94.12	94.12
10. Sense of responsibility	94.12	94.12	80.39
11. Leadership	94.12	80.39	92.16
12. Decision making	80.39	92.16	80.39
13. Coping with frustration	92.16	80.39	60.78
14. Knowledge	80.39	60.78	90.20
15. Creativeness	60.78	90.20	88.24
16. Self-reliance	90.20	88.24	68.63
17. Confidence	88.24	68.63	78.43
18. Sense of humour	68.63	78.43	70.59
19. Ability to mix well	78.43	70.59	58.82
20. Wide interest	70.59	58.82	31.37
21. Independence	58.82	31.37	23.53
22. Honesty	31.37	23.53	11.76
23. Ambition	23.53	11.76	37.25
24. Common sense	11.76	37.25	31.37
25. Logical thought	37.25	31.37	25.49

from the teachers in terms of the query as to which of the qualities the education should try to achieve were of the "Yes" or "No" type. The results of this survey in terms of the percentage of 'Yes' against each quality are given in Table 2.7. For the purpose of analysis, these qualities are then classified in terms of various priority level wise categories as defined in table 2.8.

Table 2.8

Quality Priority Level Description

Range of the percentage of the favourable i.e. 'Yes' type response	Description of the priority/ feasibility/implementation level
100% - 85%	Highest priority
84% - 70%	Good priority
69% - 55%	Average priority
54% - 40%	Less than Average priority
Less than 40%	Least priority

The results of the table 2.7 with respect to the subject matter at hand are then analysed consistent with the framework presented in table 2.8 and the emerging priority level wise categorization of the qualities that are considered to be important for being pursued by education is reported in the column 2 of table 2.9.

(III) Question 5: Results and Analysis of Qualities vis-a-vis the question as to which of these qualities the education can achieve

The responses in this context, as in the earlier situation, were also sought to be of the "Yes" or "No" type.

The details of this investigation then more or less follow those of the previous investigation.

In specific terms, the results of this survey are given in table 2.7. These results are analysed consistent with the framework reported in table 2.8 and the emerging "feasibility' level wise categorization of the qualities that can be achieved through education is reported in the column 3 of table 2.9.

(IV) Question 4: Results and Analysis of Qualities vis-a-vis the question as to which of the qualities have been achieved by the education.

The results of this investigation are given in table 2.7 These results are analysed consistent with the framework reported in table 2.8 and the emerging "implementation" level wise categorization that has been achieved through the education is reported in the fourth column of the table 2.9.

Thus, table 2.9, which is self-explanatory, presents the total analysis of the Questionnaire I.

(V) Correlations Between Responses to various Questions in Questionnaire I

The correlations between responses to various questions in questionnaire I have been given in Table 2.10.

Table 2.9

Summary of Observations emerging from Questionnaire I

Degree \ Item for which the Qualities of Questionnaire I have been analysed	Classification of Qualities for Q.1 in terms of Degree of Priority for qualities considered important for life.	Classification of qualities for Q.2 in terms of degree of priority for qualities education should achieve	Classification of Qualities from Q.3 in terms of degree of implementation for qualities education can achieve.	Classification of Qualities for Q.4 in terms of degree of feasibility for qualities education has been able to achieve
Highest	Sense of responsibility, confidence, self-reliance, common sense	Problem solving, judgement, sense of responsibility, leadership, coping with frustration, self-reliance, confidence	Diagnosis, Problem solving, Tolerance, Sense of responsibility, Decision making, Creativeness, self-reliance	Analysis, diagnosis, sympathy, tolerance, leadership, knowledge, creativeness.
Good	Judgement, Perception, understanding, logical thought, honesty, analysis, problem-solving, creativeness	Perception, Analysis, communication, understanding, sympathy, decision-making, knowledge, ability to mix well, wide interest	Perception, judgement, communication, understanding, leadership, coping with frustration, sense of humour, ability to mix well	Problem-solving, Judgement, communication, sense of responsibility, decision making, confidence, sense of humour.
Average	Diagnosis, communication, knowledge, tolerance, decision-making, ambition	Tolerance, diagnosis, creativeness, sense of humour, independence	Analysis, knowledge confidence, wide interest	Perception, coping with frustration, self-reliance, ability to mix well.
Less than Average	Independence, leadership, sympathy, wide interest, ability to mix well, sense of humour, coping with frustration.	-	Sympathy	-
Least	-	Honesty, Ambition, common sense, logical thought	Independence, honesty, ambition common sense, logical thought	Wide interest independence, honest ambition, common sense, logical thought

Table 2.10

Correlation between Responses to various Questions in Questionnaire I

Questions	Question 1: qualities that are important for life	Question 2: qualities that should be achieved by education	Question 3: qualities that can be achieved by education	Question 4: qualities that have been achieved by education
Question 1		-0.01 (slightly negative)	0.07 (slightly positive)	0.01 (slightly positive)
Question 2			0.72 (High)	0.61 (Moderate)
Question 3				0.75 (High)
Question 4				

It is against the above presented details through tables 2.1 to 2.10 that the next sub-section then summarises the emerging observations.

(VI) Summary of Observations

Table 2.9 presents the summary of observations as emerging from the analysis of the teacher responses to the Questionnaire I. Thus, as described earlier, table 2.9 gives the degree of priority-wise classification of the qualities stated under the Questionnaire I in terms of querries as to (i) firstly, which of the qualities are most necessary in life and (ii) secondly, which of the qualities should be aimed at by the higher education. Further, table 2.9 also gives the degree-of-feasibility-wise as well as the degree-of-implementation-wise classification of the various qualities stated in the Questionnaire I.

Against this, table 2.10 gives correlations between the teacher responses to the four questions of the questionnaire I. Thus, one observes that the qualities needed most in life (i.e. response to question 1) as perceived by the teacher sample studied, correlates poorly with the qualities that should be achieved by education (i.e. response to question 2). Further, the qualities needed most in life also show poor correlation with the qualities that can be achieved by education (i.e. response to question 3) as well as with the qualities that have been achieved by education (i.e. response to question 4).

Coming to the faculty response to the query as to which of the qualities should be achieved by education (i.e. response to question 2), as can be seen from table 2.10, the same shows high correlation with the qualities that can be achieved by education (i.e. response to question 3) and moderate correlation with the qualities that have been achieved by education (i.e. response to question 4). Finally, the faculty response to the qualities that can be achieved by education (i.e. response to question 3) indicate high correlation with the qualities that have been achieved by education (i.e. response to question 4

2.7.2 Results & Analysis along with emerging observations for the Questionnaire on determining objectives, Techniques and Nature of Examinations

As indicated earlier, this subsection concerns itself with results and analysis for Questionnaire II dealing with 'determination of objectives, techniques and nature of examinations' under continuous system of internal evaluation at a tertiary level as represented by BITS. Details describing this survey type questionnaire (see Appendix B) as also the type of teacher sample on which the same was administered have already been given in section 2.6.2. For the purpose of convenience, this subsection couples results and analysis with emerging observations, presenting them separately for each of the above stated three facets of examinations.

I. Results & Analysis as also emerging observations for Part A of Questionnaire II on determining Objectives of Examinations

Part A of the Questionnaire II aimed at studying objectives of evaluation components of quizzes, tests and comprehensives. Towards this, the Part A consisted of in all three questions. The first question had listed a series of educational objectives and, with reference to the courses conducted by them, for each of the above mentioned evaluation components, the teachers were requested to indicate in percentages as to how they normally weigh these objectives in the design of the evaluation components under consideration. For the purpose of the analysis, then, with reference to each of the above mentioned evaluation components, average percentage weightages based on the teacher responses were found out for each of the above indicated educational objectives. These evaluation-component-wise average-percentage-weightages for the different educational objectives are described in figure 2.1.

Coming to the second question in Part A, it aimed at exploring what percent of teachers considered that the class-room evaluation could also measure other qualities such as

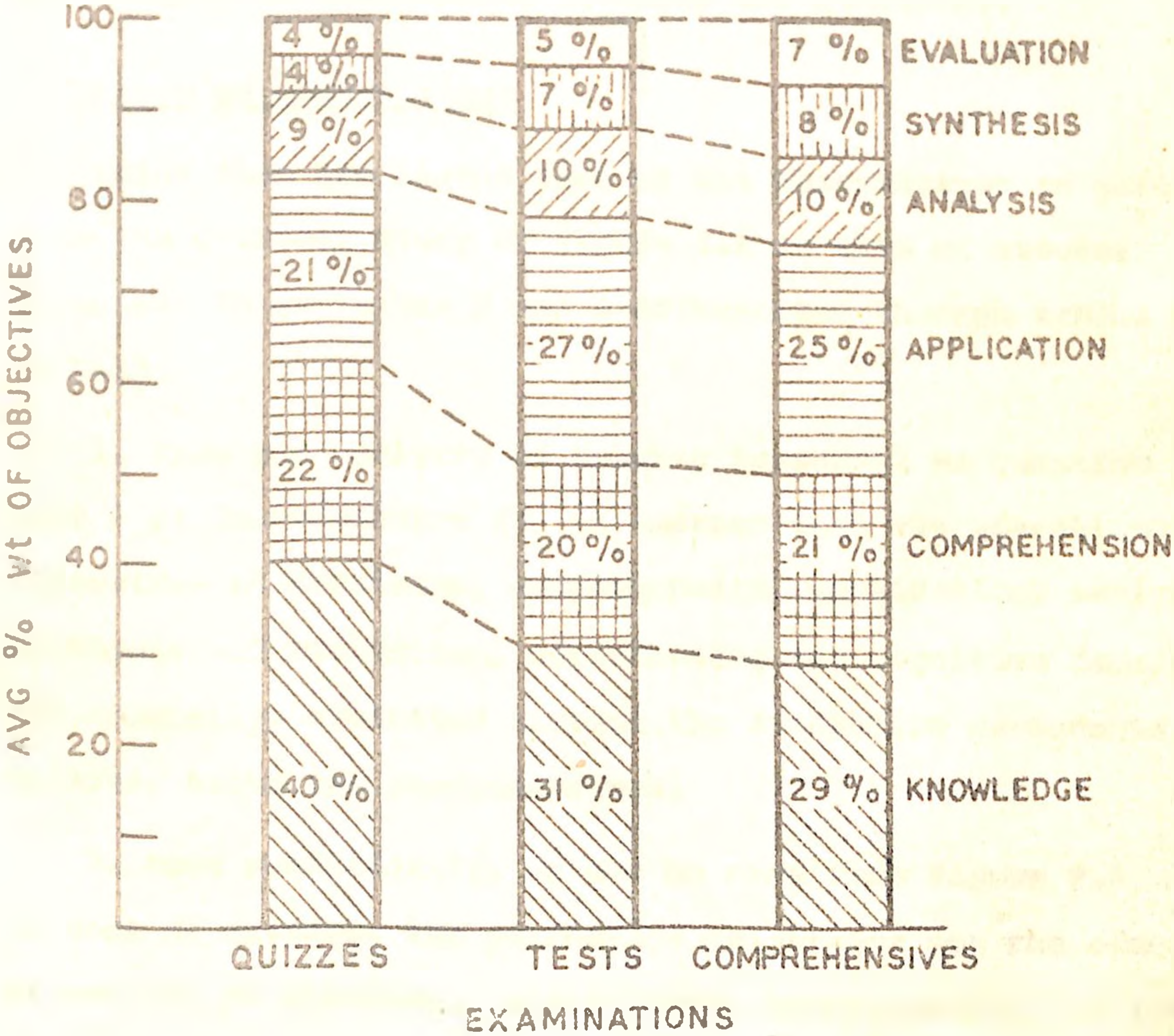


FIG. 2.1 AVERAGE PERCENTAGE OF WEIGHT GIVEN BY TEACHERS TO VARIOUS OBJECTIVES (KNOWLEDGE, COMPREHENSION ETC.) IN EXAMINATIONS.

leadership, sense of responsibility, co-operation, etc., and, from those who responded positively in the above context, the third question tried to find out what could then be the evaluation techniques for the same. Tables 2.11 to 2.13 give the analysis for results in the context of above two questions.

Summary of Observations

Below then are listed some of the observations as emrging from the critical study of figure 2.1 as also of teacher responses to questions 2 and 3 as presented through tables 2.11 to 2.13.

1. From the analysis of teacher responses to question 1 in part A of Questionnaire II, it emerges that the educational objectives of knowledge, comprehension, application, analysis, synthesis and evaluation, constituting the cognitive domain, are, normally, evaluated through the evaluation components of quizzes, tests and comprehensives.

2. More specifically, as can be seen from figure 2.1, while in case of quizzes, the percentage weightages for the educational objectives of knowledge, application, comprehension, analysis, synthesis and evaluation are 40%, 21%, 22%, 9%, 4% and 4%, respectively, the same, in case of tests, are 31%, 27%, 20%, 10%, 7% and 5%, respectively. And, in case of comprehensives, these percentage weightages turn out to be 29%, 25%, 21%, 10%, 8% and 7%, respectively.

Thus, the above indicated sequence of educational objectives also states the percentage-weightage-wise heirarchy of educational

TABLE 2.11

Analysis of Teacher Responses to Question 2 in Part A of the Questionnaire II

Discipline wise categorization of Teacher Sample	Teacher Response Frequency as per different Types of responses				
	No. of Teachers saying 'Yes'	No.of Teachers saying 'no'	No.of Teachers saying 'Not applicable'	No.of Teachers with NO response	No. of Teacher-Discipline category-wise total Teacher Response
	For each category of teacher disciplines, number in brackets show the corresponding type of the teacher response as percentage of the total teacher response for the category under consideration.				
Teachers from Professional disciplines of Engineering, M.Sc.(Tech.) inclusive of Computer Science, Instrumentation, Museum Studies, STD and Pharmacy.	10 (47.63%)	9 (42.85%)	1 (4.76%)	1 (4.76%)	21 (100%)
Teachers from Science Disciplines	19 (57.58%)	12 (36.36%)	1 (3.03%)	1 (3.03%)	33 (100%)
Teachers from the Humanities disciplines of Management, Economics, Languages and Social Studies	14 (77.78%)	4 (22.22%)			18 (100%)
Response-Type wise Total Teacher Responses	43 (59.74%)	25 (34.72%)	2 (2.77%)	2 (2.77%)	72 (100%)

TABLE 2.12

Analysis of Teacher Responses to Question 3 in Part A of the Questionnaire II seeking what evaluation-components in teachers' opinion can be used to assess qualities such as leadership, sense of responsibility, decision making, cooperation, etc.

Discipline-wise categorization of Teacher Sample	Teacher responses for various techniques of evaluation (i.e. evaluation components) as indicated by them to assess objectives listed in Question 2 in Part A of the Questionnaire II											
	Lab work Projects	Case discussions	Observation Interaction	Field work	Home project Home assignment Project work	Seminar	Viva	Question hour & discussion based classroom activity	Quiz	Regularity and behaviour	No clear cut suggestion	Teacher discipline-categorywise distribution of teacher response
Teachers from professional disciplines of Engg., M.Sc.(Tech) (inclusive of CS, Instrumentation, Museum Studies, STD) and Pharmacy.	2				7	3	1	3		1	1	18
Teachers from Science Disciplines			1		11	5	1	7	1	2	1	29
Teachers from Humanities disciplines of Management Economics, Languages & Social Studies		4	1	1	4	4	1	8		1		24
Evaluation-component-wise distribution of teacher response	2	4	2	1	22	12	3	18	1	4	2	71
Percentage-wise distribution of teacher responses	2.81	5.63	2.81	1.40	31.04	16.90	4.22	25.35	1.40	5.63	2.81	100

TABLE 2.13

Hierarchy of various student evaluation techniques as emerging on the basis of teacher responses to question 3 in part A of questionnaire II, reported in table 2.12.

Student evaluation techniques in the order of decreasing hierarchy of utility as opined by the sample of teachers studied with respect to the question of measuring qualities such as leadership, sense of responsibility, co-operation, etc.	Teacher response for the technique under consideration as % of the total teacher response for all the techniques taken together
• Home Project • Home Assignments • Project work	31.04%
• Question-hour and discussion based class-room activity	25.35%
• Seminar	16.90%
• Regularity and behaviour	5.63%
• Case discussion	5.63%
• Viva	4.22%
• Lab work/projects	2.81%
• Observation • Interaction	2.81%
• Field work	1.40%
• Quiz	1.40%

objectives as reflected through the evaluation components of quizzes, tests and comprehensives.

3. In another words, the lower order educational objectives of knowledge, comprehension and application can be seen to account for 83%, 78% and 75% of the educational-objective-wise weightages for the quizzes, tests and comprehensives, respectively. Against this, the percentage weightages for the higher order objectives of analysis, synthesis and evaluation are found to vary from 17% to 22% to 25% as one moves from quizzes to tests to comprehensives.

Thus, while on one hand it emerges that the percentage weightage for the set of lower order objectives of knowledge, comprehension and application decreases from quizzes to tests to comprehensives, on the other hand it is observed that the percentage weightage for the set of higher order objectives of analysis, synthesis and evaluation increases as one moves from quizzes to tests to comprehensives.

4. Table 2.14 given below indicates for quizzes, tests and comprehensives the degree of dominance of the set of lower order objectives of knowledge, comprehension and application over the set of higher order objectives of analysis, synthesis and evaluation.

Table 2.14

Degree of dominance of lower order objectives over the higher order objectives in the design of evaluation components of quizzes, tests and comprehensives

Evaluation component	Degree of dominance of the set of lower order objectives of knowledge, comprehension and application over the set of higher objectives of analysis, synthesis and evaluation: α = Degree of dominance for the given evaluation component $= \dfrac{\text{% weightage for the set of lower order objectives for the evaluation component under study}}{\text{% weightage for the set of higher order objectives for the evaluation component under study}}$
Quiz	$\alpha = \frac{83\%}{17\%} \simeq 5$
Tests	$\alpha = \frac{78\%}{22\%} \simeq 3.5$
Comprehensives	$\alpha = \frac{75\%}{25\%} = 3$

Thus, as can be seen from table 2.14, firstly, in the design of quizzes, tests and comprehensives, invariably, the lower order educational objectives greatly dominate the higher order objectives, and, secondly, this degree of dominance decreases from quizzes to tests to comprehensives.

5. In case of all the three evaluational components of quizzes, tests, and comprehensives, the lowest order educational objective of knowledge has highest weightage. However, consistent with the observations in (4), this percentage weightage decreases from quizzes to tests to comprehensives.

6. Thus, in summary, it is observed that, basically, the quizzes, tests and comprehensives aim at assessing the lower order qualities of knowledge, comprehension and application.

7. Table 2.11 gives the teacher-response-wise analysis for the question 2 in part A of the Questionnaire II. Thus, as can be seen from table 2.11, about 60% of the teacher sample studied have opined that the higher order objectives like leadership, sense of responsibility, decision-making, cooperation, etc. can be assessed through the class-room based evaluation.

Further, when the teacher sample studied is classified in terms of major disciplines to which the teachers belonged, it is observed that while, from the humanities disciplines of management, economics, languages and social studies, as many as 77.78% have responded favourably to the above stated query, the corresponding percentage for the group of teachers drawn from sciences is found to be 57.58%, and that for the group of teachers belonging to the spectrum of professional disciplines is observed to be only 47.63%.

Thus, while a majority percentage of teachers think that the personality traits like leadership, sense of responsibility, decision-making, etc. can be evaluated through class-room based examinations, this percentage is seen to decrease as one moves from humanities disciplines to sciences to professional disciplines.

8. Table 2.12 gives teacher-major-discipline-wise categorization of the evaluation techniques or evaluation components on the basis of teacher responses to question 3 in part A of the Questionnaire II. In specific terms, the question under consideration had aimed at seeking teachers' perception of the possible evaluation components that they would incorporate if the qualities such as leadership, decision-making, sense of responsibility, cooperation, etc. were considered measurable by them through the class-room evaluation.

Thus, as can be seen from table 2.12, a large spectrum of evaluation components emerged in the above context; namely, lab-work and lab-projects, case discussions, observation and interaction, field-work, home-projects, home-assignments and project-work, seminar, viva, question-hour and discussion based class-room activity, quiz, and, finally, regularity & student behaviour. Table 2.13 gives the hierarchy-wise ordering of the above listed evaluation-components based on their popularity with the teachers in terms of the subject matter under consideration.

Thus, as can be seen from table 2.13, the evaluation components of home-projects, home-assignments & project-work demonstrate maximum acceptability (31.04%) with teachers in terms of the task of student assessment in the context of the above indicated higher order objectives of leadership, decision-making, sense of responsibility, cooperation, etc. Next in the hierarchy, with decreasing degree of acceptability, is the evaluation component of question-hour and discussion based class-room activity, (acceptability percentage of 25.35%), followed by the evaluation component of seminar (16.9%). Against this, on the other extreme of the degree of acceptance are the evaluation components of 'lab-work & lab-projects' and 'observation & interaction' with the teacher acceptance of 2.81% and the evaluation components of 'field-work' and 'quiz' with minimum teacher acceptance of 1.4%.

Finally, it is interesting to note that while the traditional class-room based evaluation component of quiz has found minimum teacher acceptance (1.4%), the other traditional class-room based evaluation components of tests and comprehensives have failed even to get any such response from the teacher sample under consideration.

II. <u>Results & Analysis as also emerging observations for Part B of Questionnaire II on determining Techniques of Examinations</u>

Part B of Questionnaire II aimed at studying techniques of examinations/evaluation-components as adapted by teachers so as to achieve the educational objectives stated in the first question of Part A of the questionnaire under consideration. Towards this, Part B consisted of in all three questions.

The first question tried finding out how many quizzes and tests are conducted per course by teachers. Table 2.15 presents the percentage-teachers-vs-evaluation-component-frequency-wise analysis for the examinations of quizzes and tests.

TABLE 2.15

Percentage-Teachers-vs-Evaluation-Component-Frequency-wise analysis for Quizzes & Tests

Frequency / Component of evaluation	0	1	2	3	More than 3
Quiz	35%	37%	19%	1%	8%
Test		1%	7%	91%	1%

Against this, the 2nd question in part B aimed at exploring percentages of teachers incorporating evaluation components even beyond quizzes, tests and comprehensives. Various evaluation components considered in this context were home-assignments, lab-work, project-work, viva, seminar, group-discussion, etc. From the teacher sample studied, as many as 62.5% (i.e. 45 out of the total response of 72) responded favourably to the above query. The analysis of the responses of this group of teachers who responded favourably is given in Table 2.16.

Finally, the 3rd question in part B of the Questionnaire II concerned itself with finding out what educational objectives teachers keep in mind while designing the above listed evaluation components as under the 2nd question in part B. Table 2.17 gives the educational-objective-vs-evaluation-component-based analysis of the teacher responses in this context. Further, the educational objectives so emerging have been hierarchy wise ordered based

TABLE 2.16

Percentages of Teachers Incorporating various Evaluation Components extending beyond Quizzes, Tests and Comprehensives

Evaluation Component	Percentage of Teachers incorporating the evaluation component under consideration	Percentage of Teachers not incorporating the evaluation component under consideration
Home Assignments	22.22%	77.78%
Lab Work	16.67%	83.33%
Project Work	25.00%	75.00%
Viva	15.28%	84.72%
Seminar	16.67%	83.33%
Group Discussion	12.50%	87.50%
Any Other		
(i) Class Work	2.78%	97.22%
(ii) Class participation/ discussion	4.17%	95.83%
(iii) Engineering Design & Drawing	2.78%	97.22%
(iv) Lab. Test	1.39%	98.61%
(v) Field Work	1.39%	98.61%
(vi) Case-study	1.39%	98.61%
(vii) Oral presentation inclusive of debates	4.17%	95.83%

on the teacher responses for each of the educational objectives and the same (i.e. the ordering) is presented in table 2.18.

It is against the above details that below are then listed the emerging observations:

1. As can be seen from table 2.15, all the 100% of the teacher sample studied implement the evaluation component of tests. More specifically, as many as 91% of the teacher population have 3 tests per semester per course, while only 1% have more than 3 tests per semester per course. This population of 1% is observed to belong to the civil engineering group of teachers. From the remaining population of 8%, 1% teachers have only 1 test per semester per course, while the rest of the 7% have 2 tests per semester per course. This 8% of teacher population is seen to belong to the group of computer science teachers. From further critical study of the teacher responses from the computer science group, it emerges that this smaller number of tests is mainly due to the heavy emphasis placed by these teachers on the evaluation components of project-work and seminars.

   Against the above, as can be seen from table 2.15, while 8% of the sample size studied have more than 3 quizzes per semester per course, the percentage of the teacher sample conducting only 1 quiz per semester per course is as high as 37%, and as many as 35% of the population is observed not to include the evaluation component of quiz at all.

2. Coming to other details, it emerges that a majority percentage of teachers (62.5%) incorporate one or the other evaluation component in addition to those of quizzes, tests and comprehensives. Table 2.16 gives teacher-percentage wise analysis for such evaluation

components. Thus, as can be seen from table 2.16 within the frame of reference under consideration, it is the evaluation component of project-report which is most popular with the teachers (teacher percentage of 25%), followed next by home-assignments(with teacher percentage of 22.22%) and lab work and seminar (both with teacher percentage of 16.67% each). Further, the evaluation component of viva and group discussion are observed to have teacher percentage of 15.28% and 12.50%, respectively, followed, in the decreasing order of popularity, by the evaluation components of class participation/discussion and oral presentation inclusive of debates (both with teacher percentage of 4.17% each), class-work and eingineering design and drawing (with teacher percentage of 2.78% each), and, finally, lab test, field work and case study (with teacher percentage of 1.39% each).

3. Finally, table 2.17 presents various educational objectives kept in mind by teachers while designing the different evaluation components listed in (2), apart from those of quizzes, tests and comprehensives. The educational objectives so emerging have then been hierarchically ordered based on the teacher responses for each of them across the entire spectrum of evaluation components under consideration, and the same (i.e. the ordering) is reported in table 2.18.

Thus, as can be seen from table 2.18, the educational objectives of ability to make practical suggestions, ability to evaluate alternatives and problem solving ability, with the percentage-teacher-response of 21.83%, emerge as the most significant of the objectives in the above context. The above objectives, in decreasing order of significance, are then followed by the educational objectives in terms of the 'general

TABLE 2.17

Educational-Objective-vs-Evaluation-Component based analysis of Teacher Responses to Question 3 in Part B of Questionnaire II

Evaluation objectives / Evaluation Component	•Intellectual Ability •Knowledge •Comprehension	•Ability to apply knowledge to known/unknown situations •Understanding of Analytical Techniques	•Interdisciplinary Approach	•Ability to make practical suggestions •Ability to evaluate alternatives •Problem-sokving ability	•Information Processing ability inclusive of information identification, seeking, analysing and interpretation	•Ability to Articulate •Ability for written and oral presentation	•General professional personality inclusive of regularity, self reliance, sense of responsibility, leadership, sincerity, precision and reliability, etc.	No specific statement as to which objectives can be assessed.	
Home Assignments	6	5	-	6	-	5	12	2	
Lab Work	6	2	-	4	-	2	3	4	
Project Work	7	7	2	11	2	7	5	2	
Viva	4	2	-	6	-	4	6	5	
Seminar	4	4	2	9	-	8	5	1	
Group Discussion	3	4	-	4	-	3	5	2	
Any Other									
Class-work	-	-	-	-	-	-	1	-	
Class Participation & Discussions	-	1	-	-	-	-	3	-	
Engineering Design & Drawing	1	-	-	1	-	-	-	-	
Lab Test	1	1	-	-	-	-	-	-	
Field Work	-	1	-	-	-	-	-	-	
Case Study	-	1	-	2	-	-	-	-	
Oral Presentation	-	1	-	-	-	1	-	1	
Total Responses	32	29	4	43	2	30	40	17	197
Percentage-wise distribution of Teacher Responses	16.24%	14.72%	2.03%	21.83%	1.02%	15.23%	20.30%	8.63%	100%
	91.37%								

TABLE 2.18

Hierarchy of various objectives of evaluation as emerging on the basis of teacher responses to question 3 in part B of Questionnaire II, reported in Table 2.17

Evaluation objectives against the entire spectrum of techniques listed in Column 1 of table 2.16, in the order of their hierarchy as emerging from the analysis reported in table 2.16	Teacher response for the objective under considerati as percentage of the total number of teacher responses across the entire cross-section of the objectives a also the techniques.
• Ability to make practical suggestions • Ability to evaluate alternatives • Problem solving ability	21.83%
• General professional personality inclusive of regularity, self reliance, sense of responsibility, leadership, sincerity, precision & reliability, etc.	20.30%
• Intellectual Ability • Knowledge • Comprehension	16.24%
• Ability to articulate • Ability for oral and written presentation	15.23%
• Ability to apply knowledge to known/unknown situations • Understanding analytical techniques	14.72%
• Interdisciplinary approach	2.03%
• Information processing ability inclusive of information identification, seeking, analysing and interpretation	1.02%

professional personality characteristics' such as regularity, self-reliance, leadership, etc. (with percentage-teacher-response of 20.30%), which, in turn, are followed by the educational objectives of 'intellectual ability, knowledge and comprehension' (percentage-teacher-response of 16.24%).

Finally, on the other extreme of the spectrum of significance is observed the educational objective of 'interdisciplinary approach' (percentage-teacher-response of 2.03%), followed by the educational objective of information processing ability, having minimum percentage-teacher-response of 1.02%.

III. Results & Analysis as also Emerging Observations for Part C of Questionnaire II on determining the nature of Examinations

Part C of the Questionnaire II aimed at studying the nature of examinations as perceived by teachers so as to achiev the educational objectives stated in the 1st question of part A Towards this, Part C had in all 4 questions.

The first question attempted to study teachers' opinion about the type of examinations (open book/closed book, etc.) as incorporated by them in terms of quizzes, tests and comprehensives. The options to be responded were: Whether the examinations under consideration as designed by teachers were 'all or mostly open book' or 'all partly open and partly closed book' or 'all or mostly closed book'. The question was to be answered separately for quizzes, tests and comprehensives Figure 2.2 gives results for this question.

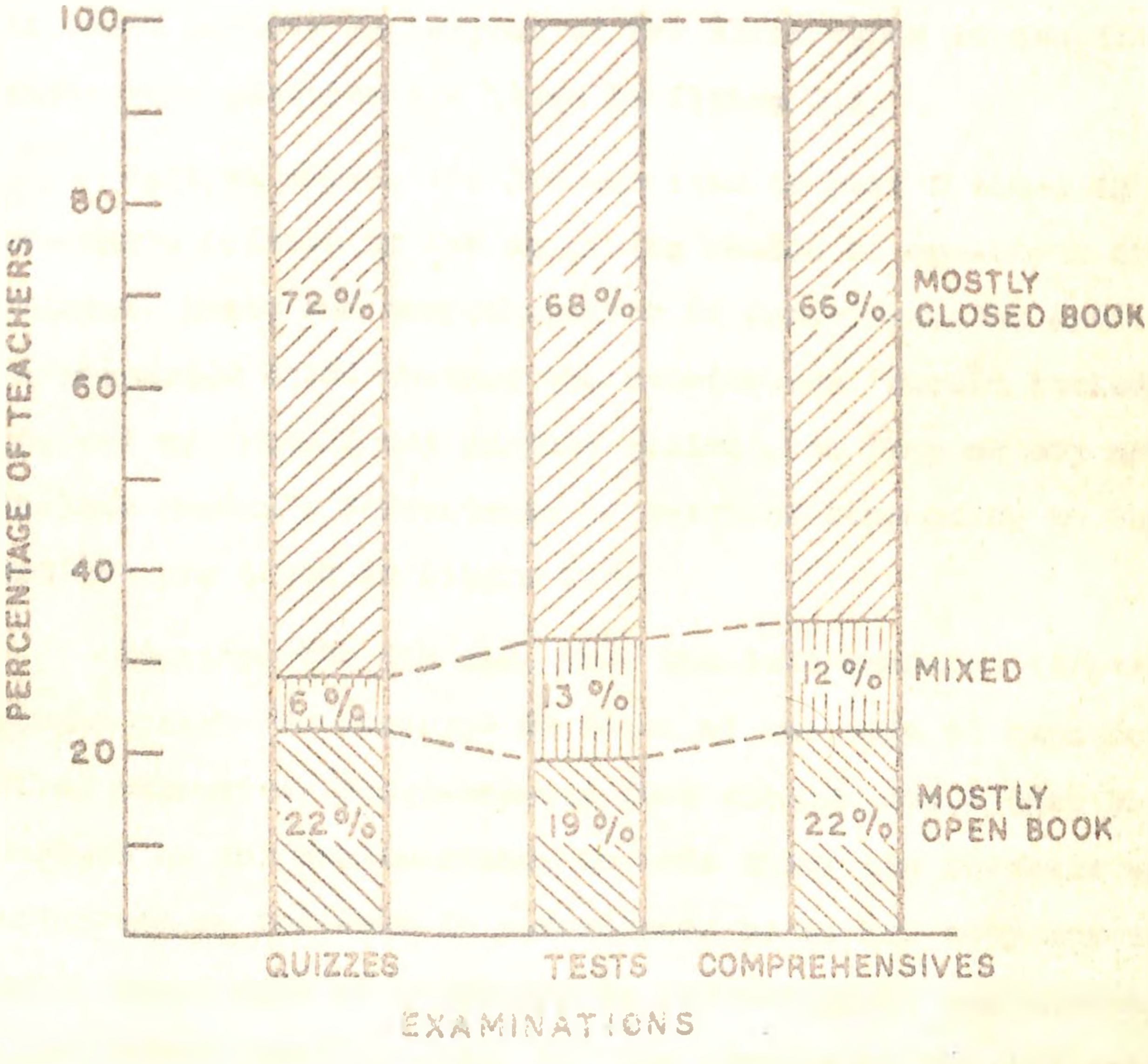


FIG. 2·2 PERCENTAGE OF TEACHERS WHO GIVE MOSTLY OPEN BOOK, MOSTLY CLOSED BOOK OR MIXED (PARTLY OPEN, PARTLY CLOSED BOOK) EXAMINATIONS.

Coming to the 2nd question, for each of the examinations of quizzes, tests and comprehensives, it aimed at exploring whether, in practice, teachers permitted choice of questions/ marks or not. Teacher percentages describing teacher responses to above options in respect of the three types of examinations under consideration are given in figure 2.3.

Against above, the 3rd question in part C aimed to know teacher's opinion as far as giving choice of questions in quizzes, tests and comprehensives is concerned. The options to be responded were: whether the examination 'should include choice' or 'should not include choice', or 'may or may not include choice'. Percentages of teachers responding to these options are given in figure 2.4.

Finally, the 4th and, thus the last question in part C sought teacher's response in terms of the type of questions (i.e. objective, short-answer, long answer, etc.) that they include in the examinations. Towards this, the teachers were requested to indicate in percentages as to how they normally weigh these type of questions in the design of the examination paper under consideration. For the purpose of the analysis, then, with reference to each of the above mentioned evaluation components (i.e. quizzes, tests and comprehensives), average percentage weightages based on the teacher responses were found out for each ~~of the~~ type of question (i.e. objective, short answer, long answer, etc.). These evaluation-component-wise average-percentage weightages for the different types of questions are given in figure 2.5.

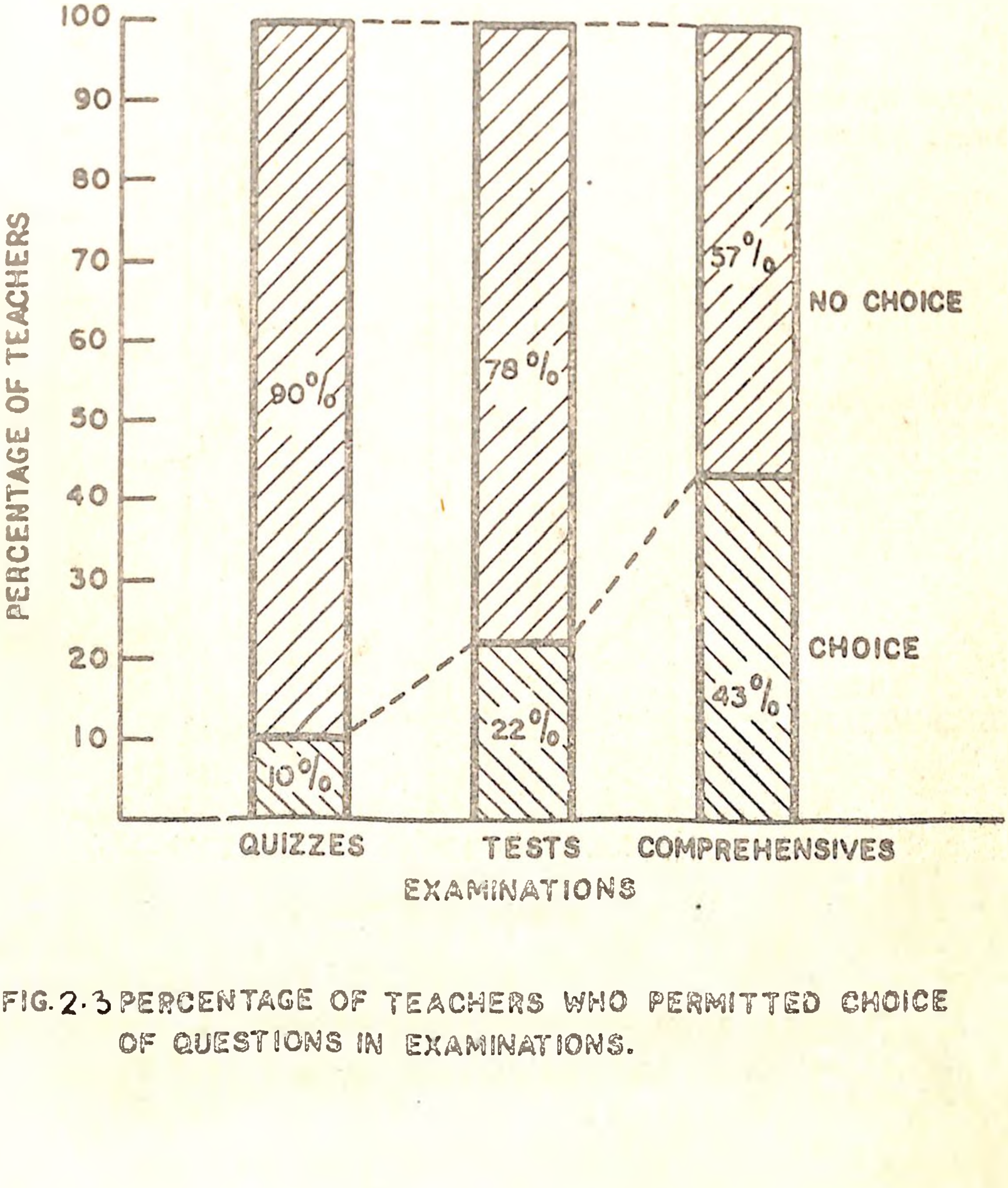


FIG. 2·3 PERCENTAGE OF TEACHERS WHO PERMITTED CHOICE OF QUESTIONS IN EXAMINATIONS.

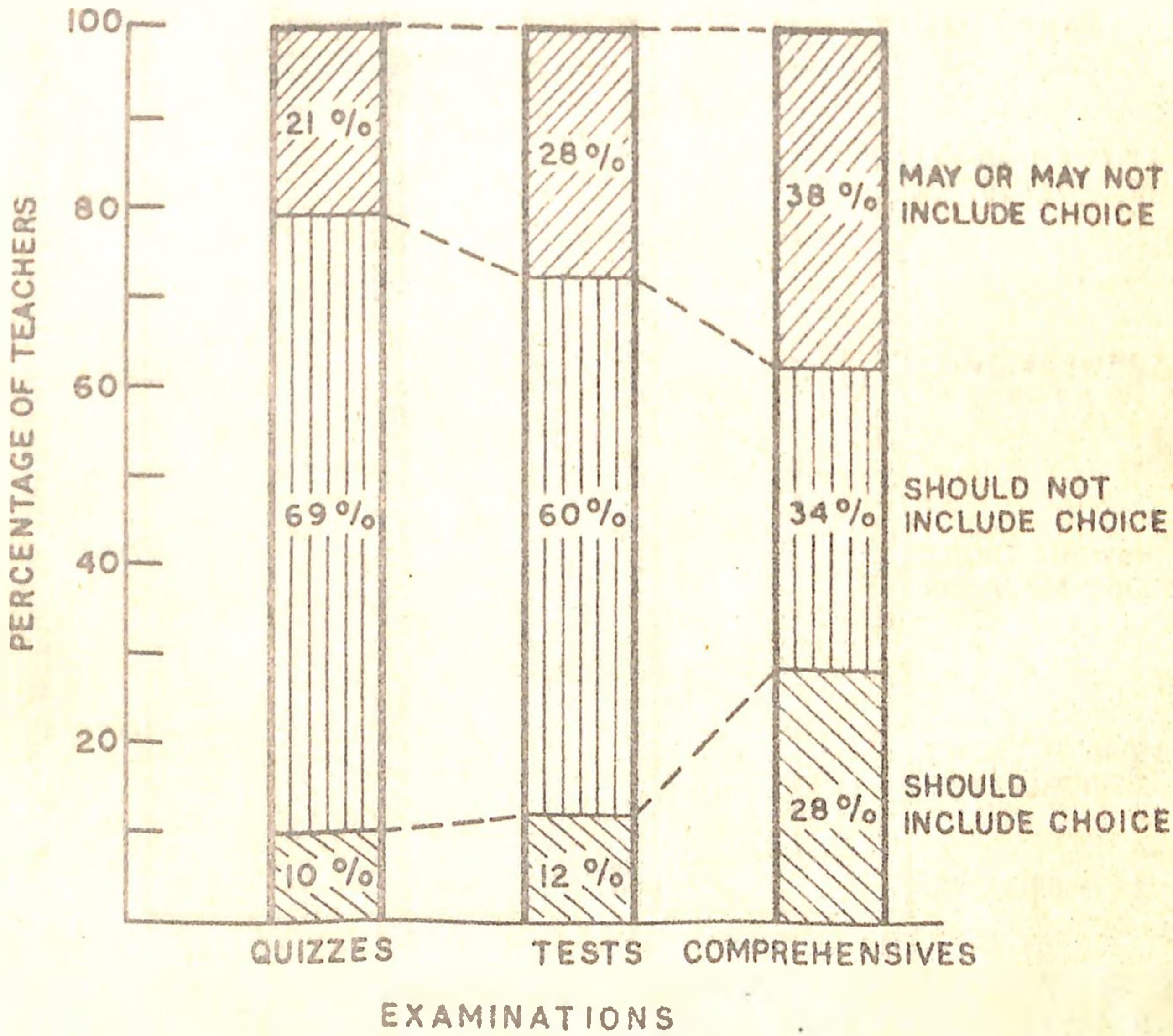


FIG. 2·4 TEACHERS' OPINION (IN TERMS OF PERCENTAGE) ABOUT GIVING CHOICE OF QUESTIONS IN EXAMINATIONS.

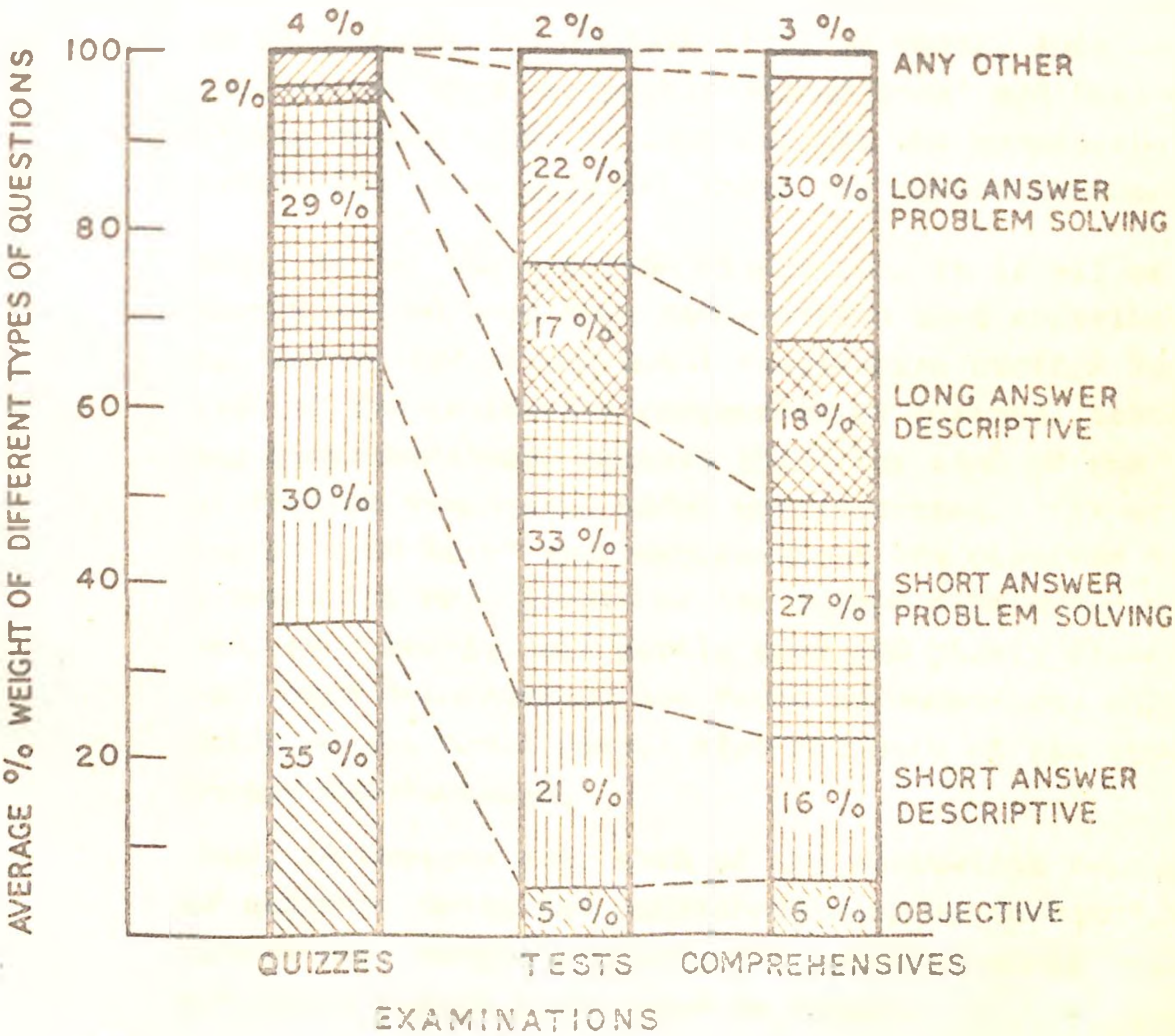


FIG. 2·5 AVERAGE PERCENTAGE WEIGHT GIVEN BY TEACHERS TO DIFFERENT TYPES OF QUESTIONS (OBJECTIVE, SHORT ANSWER ETC.) IN EXAMINATIONS.

Below are listed some of the observations as emerging from the critical study of figures from 2.2 to 2.5.

1. As far as the nature of examinations (i.e. clòsed book, open book, etc.) is concerned, the pattern of responses given to different options ('all or mostly open book', 'all partly open and partly closed book' and 'all or mostly closed book) is same for all the components of evaluation (i.e. quizzes, tests and comprehensives).

   Thus, as one can see from figure 2.2, it is all or mostly closed book type examinations that constitute between 66-72% of the total sample size studied for each of the evaluation components of quizzes, tests and comprehensives. Against this, for each of the evaluation components under consideration, 'all or mostly open book" type examinations are observed to account for only 19-22% of the sample size under study. And, as a result, all partly open and partly closed book type examinations are found to constitute only 6-12% of the total sample size for each of the above stated examinations.

   Thus, it emerges that each of the evaluation components of quizzes, tests and comprehensives, as designed by the teacher sample studied, are pre-dominantly 'all or mostly closed book' type in nature.

2. Teacher responses in terms of the practice of giving choice of questions/marks in examinations are given in figure 2.3. AS the figure indicates, whether it is a quiz, test or comprehensive, the percentage of teache 'not giving choice of questions/marks' is invariably higher than that for teachers 'giving choice of questio marks'. However, as one moves from quizzes to tests to comprehensives, the percentage of teachers 'not giving choice' decreases from 90% to 78% to 57%, respectively,

and, thus, the percentage of teachers 'giving choice' increases from 10% to 22% to 43%, respectively.

Thus, it can be concluded that while, on the whole, there is a tendency to 'not to give choice', the same is on decrease as one moves from quizzes to tests to comprehensives.

3. Figure 2.4 presents teacher-percentage-wise analysis of the teacher-responses vis-a-vis their opinions about 'giving choice in examinations'. Thus, as can be seen from figure 2.4, in case of quizzes and tests, a majority percentage of teachers (i.e. 69% and 60%, respectively), are of the opinion that the examinations should not include choice. For comprehensives, the corresponding percentage comes down to 34%. Against thi the percentage of teachers opining that 'examinations should include choice' is observed to increase from 10% to 12% to 28% as one moves from quizzes to tests to comprehensives. Finally, the percentage of teachers, who are found to be in the twilight zone, is also noted to increase from 21% to 28% to 38% as one proceeds from quizzes to tests to comprehensives.

Thus, in summary, consistent with the observations as emerging from the analysis of teacher practices vis-a-v the question of giving choices in examinations, it emerges that, on the whole, there is an increasing readiness on the part of teachers to think of 'giving of choice in examinations' as one moves from quizzes to tests to comprehensives.

4. Practices as adapted by teachers in terms of inclusion of different types of questions in quizzes, tests and comprehensives are presented in figure 2.5. As the figure indicates, quizzes mainly consist of objective (weightage 35%) and short answer (weightage 59%) type

questions. Against this, tests mainly consist of short answer (weightage 54%) and long answer (weightage 39%) type questions. A similar distribution of questions is also observed in case of comprehensives, consisting mainly of short answer (weightage 43%) and long answer (weightage 48%) type questions.

Thus, in regard to the type of questions, examinations of tests and comprehensives demonstrate a great structural similarity between each other. Further, weightage given to objective type questions is observed to vary from 35% to 5% to 6% as one proceeds from quizzes to tests to comprehensives. And, against this, the weightage given to long answer problem solving type questions is found to increase from 2% to 22% to 30% as one moves from quizzes to tests to comprehensives. Finally, while, in quizzes, the weightage given to problem solving type questions is observed to be as low as 31%, the corresponding weightages, in case of tests and comprehensives, afe found to be of the order of 55% and 57%, respectively.

This then completes the analysis of the teacher responses to the Questionnaire II, and thus, in turn, brings to close the study of the objectives of education as also of the objectives, techniques and nature of examinations as undertaken in this chapter. As indicated earlier, the next section briefly conclude the contents of this chapter.

2.8 Conclusion

This chapter has basically concerned itself with the theme of studying the teachers' perception of the objectives o education as also the objectives, techniques and nature of examinations under an internal system of continuous evaluation at a tertiary level. Towards this two questionnaires were prepared. The Questionnaire I dealt with the subject matter of the objectives of education, while the Questionnaire II was concerned with the subject of objectives, techniques and natur of examinations. The questionnaires were responded to by teachers from the entirespectrum of disciplines of engineering science and humanities. In all 51 teachers responded to the Questionnaire I, while the Questionnaire II was responded by in all 72 teachers.

Below are listed some of the main observations of the above study:

1. In the perception of teachers, the qualities of sense of responsibility, confidence, self-reliance and common sense have emerged as the ones with highest priority in the context of what the process of life needs most.
2. Against this, towards the qualities that the education should try to achieve, amongst other things, the teache have given highest priority to the qualities of problem solving(ability), judgement, leadership, self-reliance, etc.
3. In the perception of the teachers, qualities like problem solving, decision-making, self-reliance, etc. have received highest degree in terms of the ability of the education to achieve these qualities.

4. As regards to the qualities which, in the perception of the teachers, the education has been able to achieve the highest degree, amongst other things, has been given to the qualities of analysis, knowledge, leadership, etc.

5. Interestingly, while the quality of knowledge has been considered to be of only 'average' importance as far as the requirements of life are concerned, when enquired as to how far the education has been able to achieve this quality under consideration, the teachers have responded indicating 'highest' degree of achievement.

6. While common sense has been given the highest priority as far as the requirements of the life are concerned, this quality has received least priority in terms of the teachers perception vis-a-vis what the education should aim at.

7. It is interesting to note that the teacher responses to the question as to 'what qualities the life needs most' show very poor correlations with the teacher responses to the questions as to 'what education should achieve', 'what education can achieve' and 'what education has achieved'. Against this, the responses to question as to'what education should achieve' show high to moderate correlations with responses to questions as to 'what education can achieve' and 'what education has achieved', respectively. Similarly, the teacher responses to question as to 'what education can achieve' show a high correlation with the teacher responses to the question as to 'what the education has achieved'.

Thus, while there is an internal consistency between the responses to questions as to 'what education should achieve', 'what education can achieve' and 'what education has achieved', the teachers' perceptions of the qualities that the education should pursue, can pursue and has pursued are quite in isolation from their own picture of what qualities the life requires.

8. Coming to the objectives of examinations, as far as the evaluation components of quizzes, tests and comprehensives are concerned, it is observed that they are normally designed around the cognitive domain of the educational objectives

9. It is also noted that the majority of the teacher sample studied think that the higher order skills like leadership, co-operation, sense of responsibility, decision-making, etc. can be assessed through the class-room evaluation. Towards this, the evaluation components of home-projects, home-assignments and project-work are observed to be the most popular with the teachers.

10. Further, it emerges that majority of teachers from the sample studied incorporate one or other evaluation component in addition to those of quizzes, tests and comprehensives. In this context, the evaluation component of project-work is observed to be most popular with the teachers, followed by the evaluation components of home-assignments and lab-work and seminar.

11. It is observed that the educational objectives of ability to make practical suggestions, ability to evaluate alternatives and problem solving ability emerge as the most significant educational objectives in terms of the design of the various evaluation components as under (10), apart from those of quizzes, tests and comprehensives.

12. As regards to the nature of examinations, it is observed that, it is 'all or mostly closed book' type examinations that constitute between 66-72% of the total sample size studied in the above context for the examinations of quizzes, tests and comprehensives. As regards to the practice of giving choice of questions/marks, the majority of teachers do not give choice. Finally, while quizzes consist mainly of objective and short answer type questions, the tests and comprehensives include mainly the short answer and long answer type questions.

The above then is a comprehensive statement of the teachers' perception of the objectives of education as also the objectives, techniques and nature of examinations as under the internal system of continuous evaluation at the tertiary level. It is against the frame-of-reference investigated through this chapter that the teachers design various examinations, Chapter 3 studies the theme of reliability analysis for the examinations so designed.

CHAPTER 3

RELIABILITY ANALYSIS - AN AOV APPROACH

3.1 Introduction

The importance of subjecting examinations to a 'test analysis' can not be over emphasized. It is particularly vital in the case of an internal evaluation system since (i) it (test analysis) serves to establish the much-needed credibility of the system and (ii) it can provide a valuable feedback to teachers for improving upon the effectiveness of examinations which must be subjected to a rigorous scrutiny. Test analysis is used precisely for this purpose.

In more specific terms, the test analysis refers to various statistical methods which can be used to analyse the results of an examination. This chapter presents test analysis, viz., detailed statistical analysis of various examinations administered in a sample of courses at BITS. The chosen sample consists of 19 courses from various programmes and disciplines, taught at various levels across the institute. The statistical parameters which have been calculated are: reliability, coefficients of correlations between various examinations in a course and course reliability. Another important characteristic of an examination is its validity which will be discussed in the next chapter.

A brief description of the aforecited parameters is given in section 3.2 of the present chapter, whereas reliability has been discussed in some detail in section 3.3. This is followed by section 3.4 which describes the coefficient of reliability for a complete course. The review of the literature is given in section 3.5. Section 3.6 contains the objective of the present study. The chosen sample of the BITS courses is described in section 3.7 along with the method used for the collection of the data. The results, the analysis of results and the emerging observations ~~drawn~~ are presented in sections 3.8, 3.9 and 3.10, respectively.

3.2 Statistical Parameters Used in the Test Analysis

3.2.1 Coefficient of Reliability

The coefficient of reliability, denoted by R, measures the 'reliability' of an examination. Reliability of an examination refers to the consistency in different sets of marks scored if the same examination is given more than once without any intervening or additional instruction. The maximum value of R is unity which indicates that sets of marks scored in repeats of an examination are 100% consistent, i.e., the examination is 100% reliable. The other extreme R=0 would indicate a complete lack of consistency in different sets. As reliability is a very important characteristic of an examination it will be taken up for discussion in some detail in the section that follows.

3.2.2 Coefficient of Correlation

The coefficient of correlation indicates to what extent variables (tests in this case) are related, to what extent variables in one test go with variables in the other. This statistical parameter is well studied in the literature

(As can be seen in Guilford, 1978). The present study involves the use of product-moment method for the calculation of the coefficient of correlation. The method uses the relation given in the following equation (Guilford, 1978)

$$r_{XY} = \frac{xy}{N\,\sigma x\,\sigma y} \qquad ....3.1$$

where

r_{XY} = Coefficient of correlation between tests X and Y.

x = Deviation of any X score from the mean in test X.

y = Deviation of the corresponding Y score from the mean in test Y.

xy = Sum of all the products of deviations; each x deviation multiplied by its corresponding y deviation.

$\left.\begin{matrix}\sigma_x \;\& \\ \sigma_y\end{matrix}\right\}$ = Standard deviations of the distributions of X and Y scores.

3.3 Reliability of an Examination

3.3.1 Examination as a Measuring Instrument

No measuring instrument is totally free from errors. Some error or the other is always in any type of measurement, whether it is a physical or an educational measurement. The measurement error can be random or in the form of a systematic bias. The present study however is concerned with the random errors.

The random errors in measurement are never completely eliminated but efforts should be made to reduce them as much as possible. If the random errors in a measurement are sufficient small, the measurement is said to be reliable. Reliability refers to the extent to which measurements are repeatable. When the measurement is appreciably influenced by random errors, the results are not exactly repeatable because of the very nature of random errors which, being random, will be different in different sets of measurements. Generally, physical instruments like a thermometer are quite reliable. However, it is not always so in the case of educational measurements because of their qualitative nature. For example, if an achievement test administered on several different occasions gave the same results, the test could be said to be reliable. However, the achievement which was to be tested might itself not be stable as the time changes. Hence, educational measurements tend to be less reliable than physical measurements.

Examinations are instruments for measuring the ability and performance of students. It is inevitable that several kinds of errors will be inherent in various kinds of examinations. It is not possible to design a system which is totally free from errors but it is certainly essential to understand and estimate the magnitude of errors in the examinations and, if possible, establish 'limits of credibility' for examination results. Discrepancies in measurements may lead to a faulty assessment of t student's performance. Hence it is necessary to understand the different sources of errors in examinations.

The main possible sources of errors that are relevant to the internal system of evaluation can be classified as follows:

3.3.2 Errors in Examinations

(I) Subjectivity in Marking: The student's script is a major source of error. Different examiners may assign different marks to the same script. An examiner with a human mind can not be wholely free from subjectivity in marking. In addition to the contents of the answer, the examiner may be influenced by the quality of hand-writing, and the neat and tidy look of the answer book and the style of expression, etc. These factors may obviously influence the marks given to the students inspite of the usual claim that the marks are awarded only on the contents of the answer. Further, when examiners are marking a large number of answer books they would not usually mark all of them in one sitting. Hence, the varying moods and emotional and mental states of the examiners can also be one set of factors for variation in the marks awarded. Another factor which is obviously crucially important in an internal evaluation system is the examiner's personal relationship with the students and his overall opinion of the individual students. For instance, two students who have made different impressions on the examiner who has taught them in the class are likely to get different marks even if the content of their answers is identical.

(II) Content Error Due to Biased Sampling of Questions: Each examination paper represents a short sample from a large number of topics which have been taught and which form the basis for the examination. The sample may be unfair because it can not assess the entire range of the test syllabus. There is clearly a strong possibility of the sample being biased because of several reasons such as the examiners interest in particular aspects of the subject, his personal attitude towards choosing either too complex or too simple topics, etc. Heywood (1977) studies this kind of error in terms of 'suitability of the instrument or scale'. Questions which are too difficult or too easy do not help the examiner to discriminate the students' performance in a group properly. Indeed, in actual cases, then, the discrimination may be caused by the subjectivity of the examiner or by random variations.

In an internal assessment system these errors may be introduced through several ways. Since the examiner is also the teacher, during the class-room instruction students often know that he himself is going to set the question paper. Thus, he (the student) may invariably attempt to guess the possible questions in the examination from the gesture of the examiner and his emphasis on certain topics during teaching. The examiner on the other hand knows that the students are trying to guess and, therefore, depending on his own mental attitude, he may either deliberately encourage the students in guessing wrong questions or help them in guessing the right ones or even try to adopt a totally different attitude of neutrality or indifferen[ce]. Quite often, since the student knows the examiner, he may answer the questions according to the examiner's inclinations and expectations and not according to what the student really thinks should be the answer. Any such circumstances would contribute to the variation in marking which would naturally not reflect the student's true ability and achievement in the subject.

(III) Error in the Structure of a Question

Heywood (1977) calls this type of error as 'reader acuity'. If the question is ambiguous, there may be distraction in answering it properly because the students may bring their own meaning to the question .

(IV) Choice of Questions

Occasionally, the students may be given a choice of questions from a certain number of questions in the question paper (However, this practice is not very common in BITS). It is debatable whether such a choice would contribute to the error in the examination results or not, but it is obvious that this certainly would lead to the variation in test scores. Various studies (for example, Heywood 1977) have proved that the student's choice of a particular question from questions often leads to inconsistent and incomparable assessment by the examiner, depending on whether such a choice includes the examiner's favourite question or not.

(V) Error in the Test Form

The objectives of the questions in an examination should represent the objectives of the course study. Unless the examination measures what it intends to measure, it is of no use. This is called validity of the examination which will be discussed in the fourth chapter.

(VI) Error Due to the Health or Mental Condition of the Student

A student's general health and mental state can also be a source of error in the examination results. If an examination is given on a day when the student is not feeling well or is emotionally upset, his performance is likely to be worse than it would have been under normal conditions.

The errors described above contribute in varying degrees to the marks awarded to the students. It is assumed that from the point of view of the overall population these are distributed at random. The effect of such errors can be estimated by calculating the reliability of the examination which is the main objective of the work presented in the chapter. The quantitative measure of reliability is given by the coefficient of reliability as described in sub-section 3.3.4 and methods for the calculation of this coefficient are incorporated in sub-section 3.3.5.

3.3.3 <u>Safe-guards to Reduce the Errors</u>

This sub-section discusses different safeguards that can be implemented against the various categories of errors listed above. Towards this, the sub-section extensively draws from the practices followed in this context at BITS.

(I) <u>Subjectivity in Marking</u>

Students in the examination system at BITS characterized by the internal evaluation are shown their answer-books (inclusive of the comprehensive examination answer-books) after they have been corrected by the examiner. This is in a refreshing and welcome contrast to external examinations where students never get an opportunity to see their answer books after they have been checked. Hence, at BITS, in case students have any confusions or doubts, they can always have interaction with the examiners and satisfy themselves with regard to the award of marks. Such a flexibility avoids subjectivity in marking to a great extent, besides providing the necessary feedback to both the examiner and the candidate.

(II) <u>Content Error Due to Biased Sampling of Questions</u>

As pointed out earlier, unlike the traditional examination system, the evaluation at BITS is internal and continuous; and, since the evaluation is continuous, any 'biased sampling of questions' is reduced to a considerable degree. Since the students are assessed from time to time, the question papers set also constitute an authentic sample of the content taught. Further, the practice of asking more questions in a test paper can also help the examination to be less prone to this type of error.

(III) <u>Error in the Structure of the Questions</u>

Error of this type is more possible under the multisection based teaching of a course. This error can be considerably reduced by the type of the practice followed at BITS, whereby the questions are usually discussed amongst the team of instructors for the course, so as to arrive at a consensus on the content as also the structure of the questions through the necessary moderations and the revisions at the various stages of the question formulation.

(IV) <u>Choice of Questions</u>

As stated earlier, this practice can be a cause for errors in the student assessment as studied by Heywood (1977); and, therefore, may normally be discouraged. Interestingly, this practice is not very common at BITS.

(V) <u>Error in the Test Form:</u>

This arises mainly due to the discrepancy between the objectives of the course topics and those of the course examinatio

One method to reduce this kind of the error can be in terms of the practice, as followed at BITS, of issuing of course handouts at the beginning of each semester. The course handout is issued for each course of the semester, and is prepared by the instructor-in-charge of the course, who consistent with the theme of the internal evaluation, is also the examiner for the course. The handout, in specific terms, amongst other details, includes the objective of the course with reference to its description in the Institute's syllabus and also has the total examination mechanism. The handout is distributed to each of the students attending the course; thus bringing in a kind of a clarity between the examiner and the examinee in terms of the consistency between the topic and examination objectives.

(VI) Error Due to the Condition of Health or Mental State of the Student

In a system of examination characterized by only one single test/examination (normally year ending or annual), as is the situation in many of the external systems of examination, error in the assessment of the student that can creep in due to his/her improper health can indeed be of a grievious order. ~~Indeed~~ One very powerful way to avoid such a situation is to have continuous student evaluation for any course attended by him. Ofcourse, even then there can be genuine situations, in the above context, with respect to some of the tests for the course. In such case, then, the error can be reduced by opting for a practice, as followed at BITS, of "make-up" examination. A student is allowed to appear in make-up examination, if due to some 'valid' reason he is not able to appear in the examination.

3.3.4 Coefficient of Reliability

"The Reliability for any set of measurement is logically defined as the proportion of their variance that is true variance". (Guilford 1978).

Inorder to obtain a mathematical expression for the coefficient of reliability one can first write the actual score X_i obtained by the student as:

$$X_i = t_i + e_i \qquad 3.2$$

Where t_i is thetrue score of the i^{th} student and e_i is the error in his score. In a perfect system of measurement e_i would be zero and $X_i = t_i$, i.e., the true score would be the one actually obtained by the student. True score is the ideal score free from errors and can be obtained only if one has a perfectly reliable measurement.

In the above model, it is, further assumed that e_i s are distributed randomly over the whole population consisting of N students.

Thus, one, then, has the following in terms of the relation between the variances for the observed score, the true score and the error:

Let $V_o \overset{\Delta}{=}$ variance of the observed score $= \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X}_i)^2 \qquad 3.3$

Where $\overline{X}_i$ = mean of $X_i = E(X_i)$;

Let $V_t \triangleq$ Variance of the true score $= \frac{1}{N} \sum_{i=1}^{N} (t_i-\bar{t}_i)^2$ 3.4

Where $E(t_i)=\bar{t}_i$,

and $V_e \triangleq \frac{1}{N} \sum_{i=1}^{N} (e_i-\bar{e}_i)^2$ = variance of the error 3.5

where $E(e_i) = \bar{e}_i$

One, further, has:

$$E(e_i) = 0, \qquad 3.6$$

and that $E(t_i e_i) = 0$, 3.7

Then,

$$V_o = \frac{1}{N} \sum_1^N (X_i-\bar{X}_i)^2$$

$$= \frac{1}{N} \sum_1^N \left\{(t_i+e_i)-\bar{t}_i\right\}^2, \text{ as } \bar{X}_i=\bar{t}_i+\bar{e}_i; \text{ and } \bar{e}_i=0,$$

$$= \frac{1}{N} \sum_{i=1}^N \left\{(t_i-\bar{t}_i) + e_i\right\}^2$$

$$= \frac{1}{N} \sum_{i=1}^N \left\{(t_i-\bar{t}_i)^2 + 2\, e_i(t_i-\bar{t}_i)+ e_i^2\right\}$$

$$= \frac{1}{N} \sum_{i=1}^N (t_i - \bar{t}_i)^2$$

$$+ \frac{1}{N} \sum_{i=1}^N e_i\, (t_i- \bar{t}_i)$$

$$+ \frac{1}{N} \sum_{i=1}^N e_i^2 \qquad 3.8$$

But $\frac{1}{N} \sum_{i=1}^{N} e_i\, (t_i-\bar{t}_i) = E\left\{e_i(t_i-\bar{t}_i)\right\}$

$$= E(e_i t_i) - E(e_i\, \bar{t}_i)$$

$$= 0 - \bar{t}_i\, E(e_i) = 0 \qquad 3.9$$

$$\text{And, } V_e = \frac{1}{N}\sum_{i=1}^{N}(e_i-\bar{e}_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(e_i^2 - 2e_i\bar{e}_i+\bar{e}_i^2) = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \qquad 3.10$$

Therefore, substituting 3.9 and 3.10 in 3.8, we get:

$$V_o = V_t + V_e \qquad 3.11$$

The coefficient of reliability, R, defined as the ratio of true variance to observed variance, is, thus, given by

$$R = \frac{V_t}{V_o} = 1 - \frac{V_e}{V_o} \qquad 3.12$$

3.3.5 Methods for Calculating the Coefficient of Reliability

Various methods are available in the literature for calculating the coefficient of reliability. All methods have some advantages and disadvantages in terms of their own 'reliability' and convenience of application. The choice of a method would, thus, depend upon the convenience in a specific situation, on factors such as the nature of examination, the number of students and questions, the availability of data in a suitable form, etc. In this sub-section is given a brief review of some of the important methods for calculating coefficient of reliability. Further details in this regard can be obtained from many excellent books and monographs such as the Monograph on test and Item Analysis for University, (1977) by Natarajan. Some of the methods for calculating the coefficient of reliability are as follows:

I Test-retest

II Parallel form

III Split-half

IV Kuder-Richardson

V Analysis of Variance

(I) Test-retest

The test-retest method involves giving the test twice to the same group of students with an intervening time interval. Correlation between the marks obtained on the two different occasions gives the test-retest reliability coefficient. This method has the advantage of convenience and simplicity when it is easy to give a test twice to the same group of students. However, the time gap required between the two tests is a major drawback of this method. If the time gap is short, the students' performance in the second test would be influenced by their experience in the first test. If the time gap is large, their performance in the second test would differ from that in the first test due to possible changes in their intellectual level, their further readings, etc.

(II) Parallel Form

It is a modification of the test-retest method, in which the need for time gap is removed. In this method instead of administering the same test twice, two equivalent or parallel tests are given to the same population. Equivalent or parallel tests refer to the tests which have the same objectives as well as contents. The two forms of the tests are given one after the other. Theoretically parallel form seems to be a good method, but in practice it is quite difficult to make two tests which are exactly parallel.

(III) Split-half method

In this method, the reliability of a test is estimated by determining the internal consistency of the test. Unlike the two above mentioned methods, the split-half has the advantage that it requires single test administration and is quite rigorous.

In split-half method, the test is administered to a group of students in a usual manner. The performance of the students is analysed by dividing the test into two equal halves. There is no unique way of bringing about such splitting, but the usual method to split a test into two equal halves is the so called odd-even method. In this method all odd-numbered items are kept in one group and the even numbered items in the other group. An implicit assumption is that the two halves are similar or parallel. This provides the scores for each student which, when correlated, provide a measure of internal consistency of the test. The reliability of the full test can then be calculated by using Rulon's formula (Ebel,1966) as given below:

$$R = 1 - \frac{\sigma_d^{2}}{\sigma_s^{2}} \qquad 3.13$$

Where σ_d^2 is the variance of the difference between the two halves scores and σ_s^2 is the variance of the sum of the two halves scores.

In the split-half method reliability also can be calculated by using Spearman-Brown formula or a simplified formula propounded by Stanley (Ebel,1979). However, Rulon's formula given in equation 3.13 is known to be quite accurate and convenient for numerical purposes.

(IV) Kuder-Richardson Method

Like the split-half method, Kuder-Richardson (K.R.) formulae provide a measure of the internal consistency of a test by a single test administration. However, in K.R. method one need not split a test into two halves. Reliability of the test can be calculated by two alternative K.R.formulae referred as K.R. 20 and K.R. 21 (Ebel,1979) as given below:

(i) KR - 20

$$R = \frac{K}{K-1}\left[1 - \left|\frac{\sum_{i=1}^{K} p_i\, q_i}{\sigma^2}\right|\right] \qquad 3.14$$

Where K = number of items,

p_i= Proportion of responses to i^{th} item which is correct, a

q_i= Proportion of responses to the i^{th} item which is not correct.

So $(p_i + q_i)$ always equals to one.

Further, σ^2= Variance of scores on the test.

K.R.- 20 is applicable only to the tests where items are scored by giving one point if answered correctly and zero if not answered correctly. Computation by KR-20 is numerically somewhat inconvenient because one needs to know proportion of correct responses for each item. A less accurate but simpler formula to get reliability is KR 21.

(ii) KR - 21

$$R = \frac{K}{K-1}\left(1 - \frac{M(1-\frac{M}{K})}{\sigma^2}\right) \qquad 3.15$$

Where M = Mean of test scores

K = Number of items in a test, and

σ^2 = Variance of scores in a test.

K.R.- 21 gives the lower bound estimate of the reliability.

The basic disadvantage in the KR and split-half formulae is that they assume that all the items in a test are homogeneous i.e. all the items in the test are assumed to measure the same quality or characteristic and are of equal difficulty; thus putting strict quality requirements on the setting of the test paper.

(V) Analysis of Variance Approach to Reliability

Analysis of Variance(AOV) provides a convenient method to estimate the coefficient of reliability. A formula for the coefficient of reliability using the Analysis of Variance was given by Hoyt (1935). In this section a brief derivation of the formula for coefficient of reliability by AOV approach will be described.

The broad steps in the AOV approach are as follows:

1. The observed score of a student can be described through a model represented by following equations:

$$\left.\begin{aligned} x_{ij} &= t_{ij} + e_{ij} \\ \text{and} \quad t_{ij} &= \mu + \alpha_i + \beta_j \end{aligned}\right\} \quad 3.16$$

under the assumptions that:

(i) error components, e_{ij}s, occur independently and at random with zero mean and common variance, and

(ii) that they (i.e. e_{ij}s) are uncorrelated with t_{ij}s and with errors in other measurements.

Where,

(a) x_{ij} = Marks obtained by the i^{th} student in the j^{th} item.

(b) μ = Factor common to all measures (i.e. scores),

(c) α_i = Factor in the total measure signifying the contribution of the i^{th} student,

(d) β_j = Factor in the total measure signifying the contribution of the j^{th} item, and

(e) e_{ij} = Unexplained factor or the error in the measure for the i^{th} student due to the j^{th} item.

2. Then, following the same treatment as in the case of section 3.3.4 it can be shown that

$$V_o = V_t + V_e \qquad 3.17$$

Where V_o is the total variance,

V_t is the true variance,

and V_e is the error variance.

Then, the coefficient of reliability, R, is given by

$$R = \frac{V_t}{V_o} = \frac{V_o - V_e}{V_o} = 1 - \frac{V_e}{V_o} \qquad 3.18$$

It is important to note that in this model each score contains the examinee effect as well as the item effect. The present approach is thus a 'two-way' analysis of variance.

3. From a given data X_{ij}, V_o and V_e can be estimated as follows: (Guilford, 1978)

$$V_o = \frac{SE}{N-1} \quad \text{(N - Number of examinees)} \qquad 3.19$$

and

$$V_e = \frac{RSS}{(N-1)(n-1)} \quad \text{(n - Number of items)} \qquad 3.20$$

where SE represents the sum of squares for the examinees and RSS represents the remainder sum of squares. Substituting in equation 3.18 for V_o and V_e from equations 3.19 and 3.20, respectively, ~~in equation 3.18~~ yields the following expression for the coefficient of reliability

$$R = 1 - \frac{RSS}{(n-1)\ SE} \qquad 3.21$$

The sums of squares SE and RSS can be obtained as follows.

$$SE = n \sum_{i=1}^{N} \left[\frac{1}{n} \sum_{j=1}^{n} X_{ij} - \overline{X} \right]^2 \qquad 3.22$$

Where $\bar{X}$, the overall mean is defined by

$$\bar{X} = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \qquad 3.23$$

Expanding the square in equation 3.22 and using equation 3.23 gives the following equation.

$$SE = \frac{1}{n} \sum_{i=1}^{N} \left[\left(\sum_{j=1}^{n} X_{ij} \right)^2 - \frac{2}{N} \left\{ \sum_{j=1}^{n} X_{ij} \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\} + \left\{ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{N} \left\{ \sum_{j=1}^{n} X_{ij} \right\}^2 - \frac{2}{nN} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2$$

$$+ \frac{N}{nN^2} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2$$

$$= \frac{1}{n} \sum_{i=1}^{N} \left\{ \sum_{j=1}^{n} X_{ij} \right\}^2 - \frac{1}{nN} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2 \qquad 3.24$$

In a similar manner, the following equations can be obtained for S.I., the sum of squares for the items and TSS, the total sum of squares.

$$SI = N \sum_{j=1}^{n} \left[\left(\frac{1}{N} \sum_{i=1}^{N} X_{ij} \right) - \bar{X} \right]^2$$

$$= \frac{1}{N} \sum_{j=1}^{N} \left\{ \sum_{i=1}^{N} X_{ij} \right\}^2 - \frac{1}{nN} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2 \qquad 3.25$$

$$TSS = \sum_{i=1}^{N} \sum_{j=1}^{n} \left[X_{ij} - \overline{X} \right]^2$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij}^2 - \frac{1}{nN} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij} \right\}^2 \qquad 3.26$$

The remainder sum of squares RSS is obtained by using equations 3.24, 3.25 and 3.26 as follows:

$$RSS = TSS - SE - SI \qquad 3.27$$

5. Thus SE and RSS are calculated by using the above formulae to ~~which gives~~ get the coefficient of reliability in accordance with equation 3.21.

Equation 3.18 gives the coefficient of reliability obtained from the anqlysis of variance approach and is a generalised version of K.R.-20 formula (equation 3.14) which is only applicable to the objective type tests. Analysis of variance approach is most general and applicable to even essay-type tests (Harper and Misra, 1976). In general, the question papers in the BITS examinations are neither purely essay-type nor completely objective type. In such cases the use of equation 3.21 should then be the most justified.

3.3.6 Formula for Calculating Reliability of Examinations permitting choice of questions

The above formulae are not applicable to an examination in which a choice of questions is allowed. For such a case the formula given by Harper & Misra (1976) can be used. This formula is a modification of the K.R.-20 (equation 3.14) formula and is as follows:

$$R = 1 - \frac{nN \sum_{ij} X_{ij}^2 - N \sum_i (\sum_j X_{ij})^2}{(n-1)\left[N \sum_i (\sum_j X_{ij})^2 - (\sum_{ij} X_{ij})^2\right]} \qquad 3.28$$

Where

N = number of rows, i.e. the total number of examinees

n = number of columns, i.e. the total number of questions to be answered by each examinee.

X_{ij} = the marks in j^{th} question obtained by the i^{th} student.

3.4 Coefficient of Reliability for a Complete Course

The methods for calculating R as given above apply to an examination in which each question is regarded as an item. The study also aims to find out the reliability of the complete course, which at BITS generally consists of several examinations spread over a semester. To calculate the value of reliability for the complete course, one can assume a model in which each examination is treated as an item and thus use equation 3.18 based on the analysis of variance approach. In view of the general nature of the applicability of equation 3.18, this should give a reasonable estimate of reliability for the complete course. This model effectively assumes that all the examinations for the course, though carried out at different times in a semester and for different parts of the syllabus, are the components of, so to say, a 'single' examination. The value of reliability thus emerging then gives the consistency of this "single" examination, or, what can more appropriately be termed as, of the "complete" course. For the sake of brevity, in this chapter this value of Reliability is referred to as the "course reliability" (R_c).

3.5 Review of the Literature

The studies on reliability can be put under three headings:

(I) Examiner reliability, also called reader reliability

(II) Content reliability

(III) Total Reliability

(I) Examiner reliability

One of the major criticisms against examination is that the students may be wrongly classified due to error in marking. Scores assigned never give the true or real assessment of the students. This was emphasized by Taylor (1962a), Harper (1963a); George (1964) and Harper etal (1967).

In support of above observation, several other researches can be quoted. Quite a large number of studies have observed that there may be differences in marking the same set of scripts by different examiners or by the same examiner on different occasions (Mukerjee (1961), Taylor, Tluanga and Misra (1966) and Harper and Misra (1976)).

Further, Taylor (1962b) in his study "Examination of Examiners: A study of the Reliability of examination marking" observed variation in marking for equivalent set of scripts. Towards the experiment conducted in this context, he had ensured that the equivalent sets of scripts were distributed to the examiners so that the variation in marking might be attributed to the examiners and not to the scripts. In another study Taylor (1963a) studied variation in the pass percentage of equivalent set of scripts.

All the above mentioned studies, though they do not go to the extent of finding examiner reliability, are concerned wi the theme of error in marking and hence have been discussed under 'examiner reliability'. However, the researches which hav been done to study the examiner reliability are as discussed below:

Taylor, Tluanga and Misra (1966) studied the examiner reliability of English I paper of B.A. part One examination of Gauhati University. They found a correlation of 0.77.

Harper (1966a, 1967) observed the examiner reliability of 0.83 in case of a history paper of Class X of a secondary education board.

An interesting and comprehensive study on variation in marking was conducted by Harper and Misra (1976). They took a project in which 10 answer books were marked by 90 experienced examiners. This study is called 'Ninety Marking Ten'. The interesting observation is that one of the answer books which was considered the best by one experienced examiner and was given 39 (distinction) marks out of 50, was considered worst by another examiner who assigned it only 11 marks. In another case though, 77 out of 90 examiners agreed that a particular answer book was the best out of the 10, the marks awarded to it ranged from 17 to 35 out of 50; thus indicating low level of standardization in marking. In a third case the only answer book to be considered for distinction by one examiner was failed by 7. The marks awarded to it ranged from 22 percent to 76 percent.

Harper and Misra (1976) in another study, got examined for a second time the same set of 4000 answer books by the same examiners. This study is known as 'Four Thousand Remarked'. The examination selected was Class X. These 4000 answer books consisted of 1000 each from the subjects of : history, Hindi, biology and mathematics. Half of the answer books were marked a second time by the same examiner - the other half of them were marked the second time by a different examiner. The differences between the two marking of each 4000 answer books ranged from 0 to 20 marks out of 50. It is interesting to note that the geometry paper examiner differed from his own first marking by 20 points. The important conclusion observed by the study was that the marks of examiners marking their own answer books were more consisten than of those examining other examiner's answerbooks.

(II) Content Reliability

Misra (1968a) studied the content reliability of some of the matriculation examinations at Gauhati. He observed the following results:

Subject	Coefficient of reliability
English I	0.76
English II	0.63
English III	0.63
History	0.60
Mathematics	0.50
Geography	0.49

In an another study Misra (1968b) studied the content reliability of English paper of pre-university examinations of Gauhati University. He obtained the reliability coefficient of 0.80 by analysis of variance and split-half method.

Harper et al(1967) investigated the content reliability of various subjects for class X examination of a secondary board education. For mathematics paper they observed high reliabilities ranging from 0.68 to 0.90, whereas Gayen et al (1961) and Misra (1968a) found rather low reliability i.e., 0.63 and 0.50 as compared to Harper et al. It should be mentioned here that Harper et.al. studied reliability for geometry only whereas Gayen et.al and Misra did it for arithmetic, algebra and geometry, altogether.

(III) Total Reliability

The term total reliability has been coined by Harper (1976). This is the correlation between the marks given by two parallel examiners on two parallel essay tests. Misra(1976) mentions some of the studies which though not conducted under the heading 'total reliability', would be valid to be put under this category.

S.S.C.E. board Maharashtra (1960) correlated Secondary School Certificate Examination (SSCE) marks with Class X final examination for various subjects. Reliability coefficient ranged from 0.33 to 0.80 for different subjects. Gayen et.al. (1962) conducted a similar study by correlating school test examination marks with secondary board examination marks.

Harper (1966b) proved statistically that the total reliability of an essay test is equal to the product of examiner and content reliabilities. Based on Harper's approach, Misra (1968b) found the total reliability to be 0.64 for an English examination of pre-university examinations of Gauhati University.

In addition to the above it is worth mentioning some more studies which are in the direction of improving examinations. They will be discussed under following headings:

1. Question Formulation
2. Marking

1. Question Formulation

A major problem of the examination system is that of designing a good question paper. Most of the researches which have applied themselves to this aspect suggest that the questions should be made very specific and related to as directly as possible to the learning outcomes being measured. Gayen et al (1961), Hill (1967b) and Misra (1968b) found that the questions asked in the examinations are not clear. All of them suggest that the questions should be precise, so that pupils do not bring their own meaning to the questions. More ambiguous the question, more inefficient it would be in detecting the true ability of the student.

Gayen et al (1961) also pointed out that the choice of questions in an examination leads to unrealiable measure of the student's ability. Quite a number of other researches in India have also explored this problem. Harper (1962) and Taylor (1962c, 1963b), Hill (1963, 1964a & 1964b, 1965 and 1967a & 1967b), Harper (1963b) and Taylor (1962c and 1963b) suggest that choice of questions, generally, should not be allowed.

Lele et al (1962) found that students avoid difficult questions. Performance of students is found to be poorer on difficult items than on simple questions. Thus, he concluded that when choice is given students' performance depends more on the choice of questions than actually upon his knowledge of the subject matter.

Hill (1967a) analysed a mathematics question paper from a board and found that a student who knew enough about four of the eighteen topics in the syllabus got maximum marks. In another study on the general science paper, such great choice was allowed that students could safely ignore the entire physics syllabus.

Instead of long essays many authors have recommended the inclusion of compulsory, small, short answer type questions in the examinations. They hold the view that the small questions cover wide range of contents (Harper (1960), Hill (1967a, 1967b, 1967c) and Misra (1968b). Some other researches and the suggestions made by various commissions are also in favour of having objective type questions in the examinations. These recommendations are by University Education Commission (1950), Mudaliar Commission (1952), Harper (1960), Gayen et al(1961, 1962), Harper (1962), Taylor (1962d), Harper (1963b), Mukherjee (1964), Hill (1964a, 1964b, 1965), Harper (1967) and Gayen (1970).

Dave (1971) also pointed out some flaws in the question paper. He suggested following reforms:

(i) Abolition of overall options

(ii) Increasing content coverage

(iii) Increasing the proportion of short answer questions.

However, Mascarenhas (1973) in the article, 'A Model Question Paper' gives an alternate view point in the above context which instead of being based ~~in the~~ on reliability analysis uses a survey technique by involving 12 educationists. According to him, a question paper should be objective-based and include questions directly related to the objectives the curriculum is required to have and that the options should be given. Further, he suggests that the question paper should estimate the item difficulty and be so structured as to have questions ranging from easy to more difficult items.

2. <u>Marking</u>

Mahalanobis and Chakravarty (1934) in the U.P. Board examinations found significant difference in the distribution of marks for different subjects. They suggested the scaling of marks of different subjects by standard score method. Gayen et al (1961) recommended item scaling before the marks of different subjects are added. Taylor (1962e) suggested scaling for equating the marks awarded by different examiners. Baruah (1975) ~~also~~ in his monograph, 'A Hand Book of Scaling', discusses the scaling procedure. Some other important works carried out on scaling are due to Taylor (1963a), Taylor and Tluanga (1963b), Harper (1963d), Tluanga et al(1964), ~~Harper(1963),~~ Hill (1965) and Taylor, Tluanga and Misra (1966). Misra(1976) states that, "it may be observed that despite so many recommendations put forth by various experts on examination for scaling marks, no public examination body in India except the Gauhati University employs any kind of scaling in its examinations".

In this context, it may be worth mentioning the practice of normalisation followed at BITS in terms of its admission procedure. The marks obtained by each student in a public examination are normalised with respect to maximum marks awarded in that examination. For example, if a student secures x marks in a public examination and maximum marks awarded in that examination is y, then, the normalised percentage of the student having x marks will be taken to be 100 x/y. The students from different boards and public examinations are thus treated in the same footing for the purpose of admission. Thus, this normalisation enables the Institute to scale the student performances as reported by various boards, so as to take care of variations in their marking systems.

Finally, Misra (1969) in his monograph 'The effect of Randomisation and Scaling on the Errors in Examination Marks' concludes that the randomisation of scripts and scaling of marks reduce the errors in examination marks approximately by 50%.

The above, then, is a brief review of the literature on studies on reliability. As can be seen, there almost is no work on the examinations within the framework of internal evaluation system based on semester pattern. Thus, as mentioned earlier, the remaining sections of this chapter study reliability of examinations in internal continuous evaluation system.

3.6 Objective of the Present Study

The present chapter aims to study the reliability of examinations within the framework of the internal evaluation system. The system at BITS provides a very convenient data base for this research.

To begin with, in specific terms, against above objective, this chapter in its remaining sections aims to calculate the reliability of various examinations at BITS through the analysis of variance approach. As many as 83 examinations inclusive of 7 quizzes, 52 tests and 24 comprehensive examinations have been studied in this context.

Then, the sections of this chapter analyse the effect of factors such as the number of questions, number of students, level of the course, discipline of the course, etc. on the reliability of the examination.

This, then, is followed by studying the correlations between various examinations of a course, the analysis being conducted for as many as 18 courses.

Finally, the chapter devotes itself to the study of course reliability through the analysis of variance approach.

3.7 <u>Sample of Courses Chosen for Test Analysis and Data Collection Procedure</u>

~~As mentioned earlier,~~ A sample of 19 courses at BITS were chosen for the test analysis. The list of these courses is given in Table 3.1. As is apparent from the table, the select courses pertain to various disciplines and programmes across the Institute, viz., science, engineering, management and the humanities. The types and numbers various examinations analysed for ~~in~~ each course have been mentioned in Table 3.2. The total number of examinations in

TABLE 3.1

Courses Selected for Analysis

a. Core Courses

S.No.	Course No.	Course Title	Semester Academic Year	Disciplines	Level
1	SCI A111	Concepts in Science	I/1977-78	Science	I Yr.
2	PHY A112	Physics I	II/1977-78	Science	I Yr.
3	PHY A211	Modern Physics	II/1977-78	Science	II Yr.
4	MATH RA321	Probability and Statistics	II/1977-78	Science	III Yr.
5	HUM A311	Impact of Science and Technology	I/1977-78	Humanities	IIIYr.
6	HUM A311	Impact of Science and Technology	I/1978-79	Humanities	IIIYr.
7	HUM A312	Contemporary India	II/1977-78	Humanities	IIIYr.

b. Professional Courses

8	PHY A432	Solid State Physics	I/1978-79	Science	IV Yr.
9	BIO A511	Biophysics	I/1978-79	Science	V Yr.
10	MATH A561	Combinatorial Mathematics	I/1978-79	Science	V Yr.
11	ENGG A421	Control Systems Engg.	I/1978-79	Engineering	IV Yr.
12	ENGG RA512	Systems Modelling	II/1977-78	Engineering	V Yr.
13	CE A421	Soil Mechanics and Foundation Engineering	I/1978-79	Engineering	V Yr.
14	CE RA532	Water & Waste Water Treatment	I/1978-79	Engineering	V Yr.
15	EEE RA592	Introduction to Micro-prowessors	I/1978-79	Engineering	V Yr.
16	ECON A521	Developmental Planning in India	I/1978-79	Humanities	V Yr.
17	ENGG A411	Systems Analysis	I/1977-78	Management	IV Yr.
18	ENGG A411	Systems Analysis	I/1978-79	Management	IV Yr.
19	MGTS A541	Production Management	II/1977-78	Management	V Yr.

TABLE 3.2

Type and Number of Examinations Analysed in Each Course

a. Core Courses

S.No.	Course No.	No. of Quizzes	No. of Tests	No. of Comprehensives
1	SCI A111	-	3	1
2	PHY A112	-	3	1
3	PHY A211	-	-	1
4	MATH RA321	-	3	2 (Part A & B)
5	HUM A311	-	3	2 (Part A & B*)
6	HUM A311	2	3	2 (Part A & B)
7	HUM A312	1	3	1*

b. Professional Courses

8	PHY A432	2	3	1
9	BIO A511	-	3	1
10	MATH A561	-	3	1
11	ENGG A421	1(lab.Quiz)	3	1*
12	ENGG RA512	-	2	1
13	CE A421	-	3	-
14	CE RA532	1(lab.Quiz)	3	1*
15	EEE RA592	-	3	1
16	ECON A521	-	3	2 (Part A & B)
17	ENGG A411	-	2	2 (Part A & B)
18	ENGG A411	-	4 (A & B) (A & B)	2 (Part A & B)
19	MGTS A541	-	2	1

* Examinations permitting choice of questions.

As indicated earlier, the total number of examinations in these 19 courses which have been analysed consists of 7 quizzes, 52 tests and 24 comprehensive examinations.

A little clarification as regards to the number of comprehensive examinations studied would be helpful at this stage. Indeed, for 19 courses, there ought to be 19 comprehensive examinations; however, for one course, namely, - CE A421 Soil Mechanics & Foundation Engineering - data on the comprehensive examination could not be procured. Further, for six courses, comprehensive examinations had two parts each, the parts in every sense being fundamentally different in character and structure. Therefore, these have been treated as 12 examinations; thus, making the total of (13-1+12)=24 comprehensive examinations. The course numbers, course titles and the semesters of their offerings for these six courses are: HUM A311 - Impact of Science and Technology - I Semester, 1977-78; HUM A311 - Impact of Science and Technology - I Semester, 1978-79; MATH RA321 - Probability & Statistics - II Semester, 1977-78; ENGG A411 - Systems Analysis - I Semester, 1977-78; ENGG A411 - Systems Analysis - I Semester, 1978-79; and ECON A521 - Developmental Planning in India - I Semester 1978-79.

The data collected consisted of marks obtained by each student in each question of each examination. This data was collected with the help of teachers, immediately after the evaluation of answer books in each examination was completed. The teachers gave the marked answer books to the researcher before returning them to the students. Thus, the marks for each of the questions from each of the identified tests as obtained by each of the students were noted down from time to time.

Further, Table 3.1a and 3.1b , also, give the year and semester of the courses (for which the data has been analysed), as well as the level and discipline of the courses. Table 3.3 gives the number of students appearing in each examination. The number of students taking an ~~each~~ examination in a course sometimes differed from test to test because of the reasons of absence, withdrawls and cancellation of registration. In the calculation of various statistical parameters for an examination, the data has been included only in the case of those students who actually appeared in that examination.

TABLE 3.3

No. of Students Appearing in Each Examination in the Selected Courses

a. Core Courses

S.No.	Course No.	Quiz 1	Quiz 2	Test 1	Test 2	Test 3	Comprehensive
1	SCI A111	-	-	343	290	325	395
2	PHY A112	-	-	271	237	301	374
3	PHY A211	-	-	-	-	-	360
4	MATH RA321	-	-	202	143	192	231
5	HUM A311	-	-	42	59	74	94
6	HUM A311	36	43	63	47	56	70
7	HUM A312	78	-	52	57	71	84

b. Professional Courses

S.No.	Course No.	Quiz 1	Quiz 2	Test 1	Test 2	Test 3	Comprehensive
8	PHY A432	14	12	14	12	15	15
9	BIO A511	-	-	16	13	15	19
10	MATH A561	-	-	16	17	12	17
11	ENGG A421	56(lab.)	-	131	80	122	136
12	ENGG RA512	-	-	-	121	59	181
13	CE A421	-	-	12	21	26	-
14	CE RA532	12(lab.)	-	14	13	12	14
15	EEE RA592	-	-	50	34	47	48
16	ECON A521	-	-	25	18	18	24 Part A 25 Part B
17	ENGG A411	-	-	-	53	37	74
18	ENGG A411	-	-	55	-	21	70
19	MGTS A541	-	-	30	20	-	26

3.8 Results

This section, consistent with the objectives stated in section 3.6, presents the results on the coefficient of reliability for the 19 selected courses. The titles of these 19 courses are same as those listed in Table 3.1. These results are given in part a's of tables from 3.4 to 3.22.

Then this section proceeds to give the results on the coefficients of correlations between various examinations for the courses listed in Table 3.1, except for the course CE A421 - Soil Mechanics & Foundation Engineering. These results are given in part b's of Tables 3.4 to 3.22.

Finally, this section gives in its Table 3.23 results on course reliability for the courses listed in Table 3.1, excepting two courses, namely, MATH RA321 - Probability & Statistics and CE A421 - Soil Mechanics & Foundation Engineering.

It may be mentioned that the courses CE A421 - Soil Mechanics & Foundation Engineering and MATH RA321 - Probability & Statistics were, as indicated above, excluded from calculations for the reasons of the data not being available in the proper form.

Coming to other details, it may be mentioned that all above calculations were carried out on IBM 1130 computer at BITS using Fortran programmes and the same have been listed in appendix I (i to iv).

TABLE 3.4

1. Course - SCI A111 - Concepts in Science

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	17	15	0.62
Test 2	14	15	0.58
Test 3	19	15	0.61
Comprehensive	32	35	0.67

b. Correlation coefficients betweem pairs of examinations

	Quiz 2	Test 1	Test 2	Test 3	Project report	Comprehensive
Quiz 1	0.24	0.37	0.43	0.42	0.03	0.41
Quiz 2		0.36	0.34	0.33	0.07	0.39
Test 1			0.54	0.52	0.01	0.60
Test 2				0.52	0.06	0.63
Test 3					0.08	0.60
Project report						0.12
Comprehensive						

*In addition, 2 quizzes of 5 marks each and a project report of 10 marks were also given.

TABLE 3.5

2. Course - PHY RA112 - Physics I

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	12	18	0.60
Test 2	10	18	0.59
Test 3	8	16	0.65
Comprehensive	7	35	0.70

b. Correlations coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Compre.
Quiz 3	0.47	0.46	0.37	0.52
Test 1		0.52	0.38	0.47
Test 2			0.47	0.51
Test 3				0.56
Comprehensive				

*In addition, a quiz of 11 marks was also held.

TABLE 3.6

3. Course - PHY A211 - Modern Physics

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Comprehensive	15	30	0.91

b. Correlation coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Lab work	Compre.
Lab quiz	0.43	0.49	0.48	0.46	0.50
Test 1		0.52	0.47	0.23	0.58
Test 2			0.58	0.33	0.60
Test 3				0.48	0.60
Lab work					0.32
Comprehensive					

*In addition, three tests of 15 marks each, a quiz of 10 marks and the lab work of 15 marks were also given.

TABLE 3.7

4. Course MATH RA321 - Probability & Statistics

a. Reliability of examinations

Examinations*	No. of questions	Actual weight	Reliability Coefficients
Test 1	6	30	0.59
Test 2	6	30	0.56
Test 3	6	30	0.73
Comprehensive A	16	25	0.73
Comprehensive B	14	50	0.82

b. Correlation Coefficients between pairs of Examinations

	Test 1	Test 2	Test 3	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Comp hens
Test 1		0.39	0.44	0.17	0.22	0.19	0.22	0.34	0.1
Test 2			0.50	0.28	0.30	0.30	0.32	0.31	0.1
Test 3				0.095	0.25	0.21	0.30	0.35	0.[illegible]
Quiz 1					0.23	0.18	0.34	0.24	0.[illegible]
Quiz 2						0.13	0.27	0.29	0.[illegible]
Quiz 3							0.19	0.17	0.0
Quiz 4								0.40	0.1
Quiz 5									0.1
Compre-hensive									

* In addition, five quizzes of 7 marks each were also held.

TABLE 3.8

5. Course - HUM A311 - Impact of Science & Technology

(I Semester 1977-78)

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	19	20	0.70
Test 2	11	20	0.46
Test 3	4	20	0.45
Comprehensive*** Part A	13	15	0.61
Comprehensive** Part B	3 out of 4**	15	0.61

b. Correlation coefficients between pairs of examinations

	Quiz 2	Test 1	Test 2	Test 3	Compre.
Quiz 1	0.18	0.25	0.27	0.48	0.37
Quiz 2		0.40	0.16	0.33	0.42
Test 1			0.23	0.37	0.57
Test 2				0.39	0.35
Test 3					0.50
Comprehensive					

*In addition, 2 quizzes of 5 marks each were also held.

**Comprehensive examination Part B permitted choice of questions.

***The question paper of the comprehensive examination contained the parts A and B. These parts were quite different in nature hence their reliability coefficients were calculated separately.

TABLE 3.9

6. Course - HUM A311 - Impact of Science & Technology (I Semester 1978-79)

a. Reliability of examinations

Examinations	No. of questions	Percentage weight	Reliability coefficient
Quiz 1	6	5	0.75
Quiz 2	8	5	0.44
Test 1	12	20	0.47
Test 2	8	20	0.59
Test 3	5	20	0.42
Comprehensive* Part A	16	18	0.77
Comprehensive* Part B	2	12	0.48

b. Correlation coefficients between pairs of examinations

	Quiz 2	Test 1	Test 2	Test 3	Compre.
Quiz 1	0.22	0.47	0.45	0.40	0.06
Quiz 2		0.54	0.54	0.37	0.47
Test 1			0.60	0.53	0.28
Test 2				0.59	0.37
Test 3					0.17
Comprehensive					

*The question paper of the comprehensive examination contained two parts A and B. These parts were quite different in nature hence their reliability coefficients were calculated separately.

TABLE 3.10

7. Course - HUM RA312 - Contemporary India

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Quiz	12	5	0.77
Test 1	2	15	0.42
Test 2	3	15	0.09
Test 3	3	15	0.39
Comprehensive	4 out of 5**	40	0.56

b. Correlation coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Home Assign-ment	Compre.
Quiz	0.36	0.41	0.29	0.32	0.37
Test 1		0.37	0.37	0.38	0.40
Test 2			0.42	0.34	0.33
Test 3				0.51	0.70
Home assign-ment					0.64
Comprehensive					

*In addition, home assignment of 10 marks was also given.

**Examination permitted choice of questions.

TABLE 3.11

8. Course - PHY A432 - Solid State Physics

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Quiz 1	9	5	0.40
Quiz 2	7	5	0.73
Test 1	4	15	0.24
Test 2	4	15	0.74
Test 3	3	15	0.19
Comprehensive	11	30	0.40

b. Correlation coefficients between pairs of examinations

	Quiz 2	Quiz 3	Test 1	Test 2	Test 3	Home Assign-ment	Compre.
Quiz 1	0.08	0.60	0.66	0.28	0.39	0.45	0.53
Quiz 2		0.36	0.06	0.64	0.48	0.38	0.42
Quiz 3			0.34	0.48	0.21	0.64	0.39
Test 1				0.44	0.35	0.39	0.46
Test 2					0.78	0.51	0.67
Test 3						0.22	0.80
Home assign-ment							0.54
Compre-hensive							

*In addition, quiz 3 of 5 marks and a home assignment of 10 marks were also given.

TABLE 3.12

9. Course - BIO A511 - Biophysics

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	3	15	0.39
Test 2	3	15	0.22
Test 3	3	15	0.27
Comprehensive	19	45	0.83

b. Correlation coefficients between pairs of examinations

	Quiz 2	Test 1	Test 2	Test 3	Compre.
Quiz 1	0.33	0.48	0.50	0.58	0.72
Quiz 2		0.32	0.60	0.12	0.56
Test 1			0.32	0.56	0.51
Test 2				0.50	0.69
Test 3					0.68
Comprehensive					

*In addition, quiz 1 and quiz 2 of 5 marks each were also held.

TABLE 3.13

10. Course - MATH A561 - Combinatorial Mathematics

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	6	15	0.48
Test 2	4	15	0.10
Test 3	5	15	0.66
Comprehensive	7	30	0.67

b. Correlation coefficients between pairs of examinations

	Test 2	Test 3	Seminar	Compre.
Test 1	0.76	0.74	0.73	0.45
Test 2		0.65	0.42	0.33
Test 3			0.69	0.76
Seminar				0.58
Comprehensive				

*In addition, a seminar of 25 marks was also held.

TABLE 3.14

11. Course - ENGG A421 - Control Systems Engineering

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Lab quiz	10	10	0.47
Test 1	3	15	0.52
Test 2	5	15	0.41
Test 3	4	15	0.48
Comprehensive	5 out of 6**	35	0.68

b. Correlation coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Lab report	Comprehensive
Lab quiz	0.29	0.27	0.38	0.45	0.37
Test 1		0.29	0.38	0.44	0.37
Test 2			0.53	0.49	0.61
Test 3				0.55	0.68
Lab report					0.60
Comprehensive					

*In addition, a lab report of 10 marks was also given.

**Examination permitted choice of questions.

TABLE 3.15

12. Course - ENGG RA512 - Systems Modelling

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 2	2	18	0.18
Test 3	4	18	0.52
Comprehensive	6	36	0.63

b. Correlation coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Compre.
Quiz 3	0.33	0.34	0.40	0.41
Test 1		0.44	0.34	0.59
Test 2			0.38	0.45
Test 3				0.46
Comprehensive				

*In addition, test 1 of 18 marks and a quiz of 10 marks were also held.

TABLE 3.16

13. <u>Course - CE A421 - Soil Mechanics & Foundation Engineering</u>

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	4	20	0.26
Test 2	5	20	0.26
Test 3	3	20	0.42

*In addition, a quiz of 5 marks and the comprehensive of 35 marks were also held.

Note :- Correlation coefficient between various examinations in this course could not be calculated due to inadequate data.

TABLE 3.17

14. Course - CE RA532 - Water & Waste water Treatment

a. Reliability of examinations

Examinations	No. of questions	Percentage weight	Reliability coefficient
Lab quiz	8	20	0.92
Test 1	16	15	0.38
Test 2	8	15	0.14
Test 3	7	15	0.38
Comprehensive	5 out of 7*	35	0.35

b. Correlation coefficients between pairs of examinations

	Test 1	Test 2	Test 3	Compre.
Lab quiz	0.16	0.91	0.68	0.81
Test 1		0.17	-0.11	0.05
Test 2			0.77	0.89
Test 3				0.87
Comprehensive				

*Examination permitted choice of questions.

TABLE 3.18

15. Course - EEE RA592 - Introduction to Microprocessors

a. Reliability of examinations

Examinations	No. of questions	Percentage weight	Reliability coefficient
Test 1	3	20	0.58
Test 2	5	20	0.42
Test 3	11	20	0.93
Comprehensive	5	40	0.50

b. Correlation coefficients between pairs of examinations

	Test 2	Test 3	Comprehensive
Test 1	0.44	0.48	0.34
Test 2		0.57	0.52
Test 3			0.59
Comprehensive			

TABLE 3.19

16. Course - ECON A521 - Developmental Planning in India

a. Reliability of examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 1	10	10	0.70
Test 2	15	17	0.79
Test 3	10	18	0.47
Comprehensive Part A**	14	14	0.66
Comprehensive Part B**	4	21	0.57

b. Correlation coefficients between pairs of examinations

	Quiz 2	Test 1	Test 2	Test 3	Comprehensive
Quiz 1	0.01	0.39	0.51	0.53	0.31
Quiz 2		-0.10	-0.14	0.10	0.54
Test 1			0.61	0.63	0.55
Test 2				0.68	0.47
Test 3					0.59
Compre-hensive					

*In addition, quiz 1 and quiz 2 of 10 marks each were also held.

**The question paper of the comprehensive examination contained two parts - A & B. These parts were quite different in nature hence their reliability coefficients were calculated separately.

TABLE 3.20

17. Course - ENGG A411 - SYSTEMS ANALYSIS

(I Semester 1978-79)

a. Reliability of Examinations

Examination*	No. of questions	Percentage weight	Reliability Coefficient
Test 1 Part A	12	8.5	0.57
Test 1 Part B	4	9	0.04
Test 3 Part A	11	8.5	0.40
Test 3 Part B	2	9	0.58
Comprehensive Part A	23	15	0.78
Comprehensive Part B	6	25	0.64

b. Correlation Coefficients Between Pairs of Examinations

	Test 2	Test 3	Quiz	Comprehensive
Test 1	0.53	0.58	0.49	0.55
Test 2		0.49	0.56	0.57
Test 3			0.52	0.62
Quiz				0.59

* In addition, Test 2 of 17.5 marks and a quiz of 7.5 marks were also held.

TABLE 3.21

18. Course - ENGG A411 - Systems Analysis
(I Semester, 1977-78)

a. Reliability of Examinations

Examinations*	No. of questions	Percentage weight	Reliability coefficient
Test 2 Part A	14	6	0.73
Test 3 Part A	12	6	0.66
Comprehensive Part A	27	12.5	0.67
Comprehensive Part B	3	25	0.44

b. Correlation Coefficients Between Pairs of Examinations

	Test 2	Test 3	Quiz	Project Report	Comprehensive Part B	Comprehensive Part A
Test 1	0.62	0.52	0.37	0.42	0.62	0.65
Test 2		0.65	0.43	0.66	0.65	0.70
Test 3			0.43	0.53	0.66	0.65
Quiz				0.39	0.49	0.39
Project Report					0.52	0.60
Comprehensive Part B						0.69
Comprehensive Part A						

* In addition, test 1 or 15 marks, test 2 and 3 Part B of 9 marks each, a quiz of 5 marks and a project report of 12.5 marks were also given.

TABLE 3.22

19. Course - MGTS A541 - Production Management II

a. Reliability of examinations

Examinations	No. of questions	Percentage weight	Reliability coefficient
Test 1	12	15	0.67
Test 2	8	15	0.69
Comprehensive	16	30	0.39

b. Correlation coefficients between pairs of examinations

	Test 2	Report 1	Report 2	Report 3	Report 4	Asgnt.1 1	Asgnt. 2	Compr
Test 1	0.42	0.47	0.47	0.28	0.48	0.39	0.29	0.41
Test 2		-0.09	0.57	0.11	0.50	0.42	0.36	0.25
Report 1			0.18	0.22	0.21	0.28	0.05	0.17
Report 2				0.17	0.49	0.36	0.28	-0.01
Report 3					0.56	0.18	0.17	0.28
Report 4						0.43	0.60	0.45
Assignment 1							0.69	0.10
Assignment 2								-0.04
Comprehensive								

*In addition, 4 reports of 5, 8, 7 & 10 marks respectively and 2 assignments of 5 marks each were also held.

TABLE 3.23

Course Reliability of the selected courses

Sl. No.	Course Number	Course Title	No.of stu-dents	No. of exami-nations	Course relia-bility
1.	SCI A111	Concepts in Science	396	7	0.75
2.	PHY A112	Physics I	363	5	0.78
3.	PHY A211	Modern Physics	363	6	0.82
4.	HUM A311	Impact of Sci. & Tech. (I Sem. 1977-78)	95	6	0.73
5.	HUM A311	Impact of Sci. & Tech. (I Sem. 1978-79)	61	6	0.66
6.	HUM A312	Contemporary India	83	6	0.75
7.	PHY A432	Solid State Physics	15	8	0.85
8.	BIO A511	Biophysics	19	6	0.73
9.	MATH A561	Combinatorial Mathematics	17	5	0.87
10.	ENGG A421	Control Systems Engineering	136	6	0.75
11.	ENGG RA512	Systems Modelling	176	5	0.76
12.	CE RA532	Water & Waste Water Treatment	15	5	0.82
13.	EEE RA592	Introduction to Microprocessors	49	4	0.78
14.	ECON A521	Developmental Planning in India	25	6	0.73
15.	ENGG A411	Systems Analysis (I Sem 1977-78)	76	7	0.88
16.	ENGG A411	Systems Analysis (I Sem 1978-79)	71	5	0.76
17.	MGTS A541	Production Management	26	9	0.78

3.9 Analysis of Results

In this section the coefficients of reliability would be analysed first. Next, the study would analyse the coefficients of correlations between various examinations of a course. Finally, the analysis would be done for the results on course reliability.

3.9.1 Analysis of Results on the Reliability of Examinations:

Traditionally, any effort in the direction of the test analysis, invariably, attempts to answer questions like, how the test-reliability is related to the items incorporated in the test-paper, how does the test reliability for the courses from, say, sciences compare with the same for the courses from the disciplines of the humanities, how reliable are the objective type tests as compared to the essay type tests and so on.

It is against this backdrop, that this sub-section, to begin with, proposes to analyse the results on the reliability of examinations presented in the previous section in terms of parameters such as:

(i) Number of questions,
(ii) Duration of examination,
(iii) Number of students,
(iv) Level of the course,
(v) Core courses versus professional courses,
(vi) Discipline of the course,
(vii) Type of examination (short answer, long answer, etc.), and
(viii) Choice given in an examination.

Towards the above, based on the type of questions in examination and based on the examination duration, the set of examinations under consideration can be classified into three main categories - quizzes, tests and comprehensives. Thus, while on the one hand this chapter analyses the data base and the emerging results as available from the 83 examinations at the macro-level taking the examinations all together, on the other hand it also analyses the results at the micro-level taking each of the above stated categories separately.

Further, in order to examine how the reliability of an examination is affected by the aforementioned eight parameters, the examinations under the above mentioned categories have been considered in two groups according to their coefficients of reliability. For the quizzes the groups chosen are:

(a) those with the coefficient of reliability less than 0.4, and

(b) those with the coefficient of reliability equal to or more than 0.4;

thus, making the threshold value for the coefficient of reliability for quizzes to be 0.4. For the tests, the corresponding threshold value of reliability is taken to be 0.5 and for the comprehensives, the same is 0.6. For all the examinations taken together, this value emerges to be 0.5. The choice of these demarcating values of the coefficients of reliability for different categories was dictated by the observed clustering of the reliability values of examinations on appropriate graphs plotting the coefficients of reliability against the appropriately selected parameters, one of them being the number of questions in an examination (for illustration see figures 3.3a, 3.3b, 3.21a and 3.21b).

It is against above framework that below is presented the analysis of the reliability of examinations for each of the earlier listed parameters.

(i) Reliability Observations According to the Number of Questions

Here the analysis of results has been done in following manner:

1. Macro-level Analysis: For this, all the examinations, irrespective of their categories, have been analysed together as a set.

2. Micro-level Analysis: Here, individual categories of examinations such as quizzes, tests and the comprehensive examinations have been analysed separately.

The total number of examinations that has been studied is 83. These consist of 7 quizzes, 52 tests and 24 comprehensives.

Macro-level Analysis

(a) Analysis for all the examinations taken together

To study the relationship between the reliability coefficients and the number of questions in an examination, to begin with, the entire spectrum of the 83 examinations was classified in terms of the number of questions in the examination. Thus, depending on the number, representing the total number of questions in an examination, there emerged one or more than one reliability coefficient(s). From here was, then, calculated an average reliability coefficient for each number representing the number of questions in an examination. This information is given in the graph (3.1) showing the variation of the average reliability coefficients against the total number of questions in an examination.

Earlier this sub-section ~~Section 3.9.1~~ has mentioned about the threshold reliability coefficients as obtained at the micro-level for various categories of quizzes, tests and comprehensives as also at the macro level for all the examinations taken together. As indicated earlier, these thresholds were obtained while considering the clusters in the graph showing the reliability coefficients as the function of parameters like the ~~the~~ number of questions in an examination. Understandably, this data-base had also afforded the researcher an opportunity to analyse how the examinations were placed in terms of the number of questions, that is, to say, what percentage of total examinations had number of questions in an examination, say, equal to or more than certain number say x_1 or what percentage had number of questions, say, less than x_2 or, say, how many examinations had number of questions between x_1 and x_2 and so on. When the data base thus was studied in this context, clusters emerged classifying examinations in different groups in terms of the number of questions in an examination or, in other words, threshold values of number of questions in an examination also emerged for different

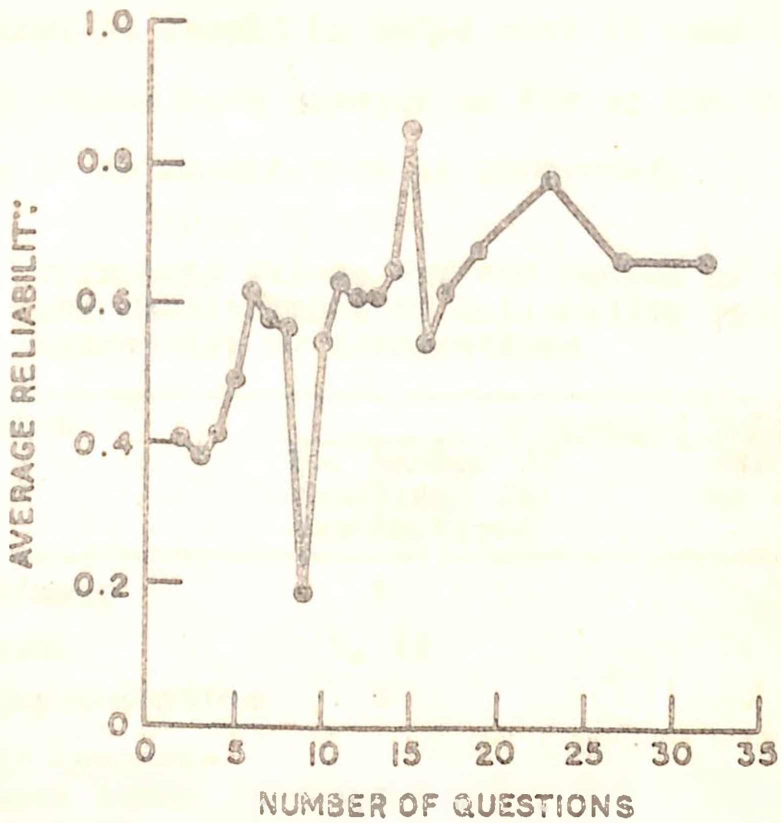


FIG. 3·1 VARIATION OF AVERAGE RELIABILITY WITH NUMBER OF QUESTIONS IN THE EXAMINATIONS (QUIZZES, TESTS AND COMPREHENSIVES).

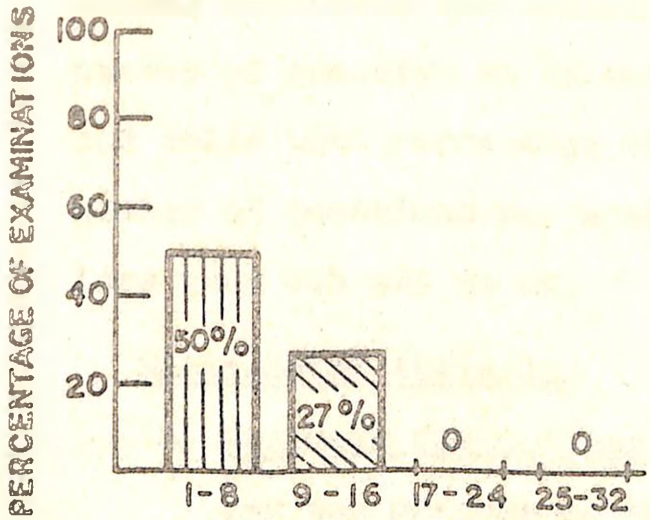


FIG. 3·2 PERCENTAGE OF EXAMINATIONS (QUIZZES, TES 'S AND COMPRE) WITH R<0.5.

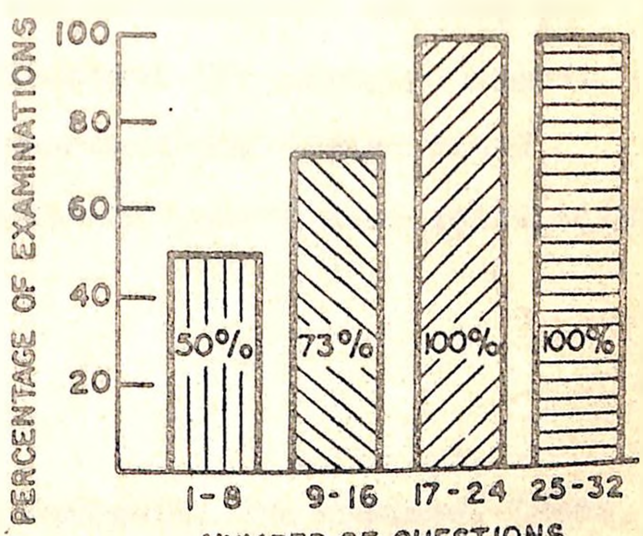


FIG. 3·3 PERCENTAGE OF EXAMINATIONS (QUIZZES, TESTS AND COMPRE) WITH R ≥0.5.

(For illustration see figures 3.3a and 3.3b)
categories. These threshold values can be seen from the table given below. It should be noted that in case of tests, two threshold values have emerged as far as the number of questions in an examination is concerned.

Table 3.23a

Threshold Values for the Number of Questions and Coefficients of Reliability for various Categories of Examinations

Categories		Threshold Values	
		For Number of questions in examinations	For coefficient of Reliability
Micro-level	Quizzes	7	0.4
	Tests	7, 13	0.5
	Comprehensives	7	0.6
Macro-level	All examinations taken together	7	0.5

Figures 3.2 and 3.3 give how the examinations are placed in terms of their relationship with the threshold values vis-a-vis the reliability coefficients as also the number of question in an examination; for example, figure 3.2 tells what percentage of examinations having total number of questions per examinations between 1-8 have reliability less than 0.5 and so on.

Micro-level Analysis

(b) Analysis for Quizzes

For the purpose of the analysis, the quizzes, based on the threshold value for the number of questions in an examination, were grouped into following categories:

(i) Quizzes where the number of questions is 7 or less than 7, and

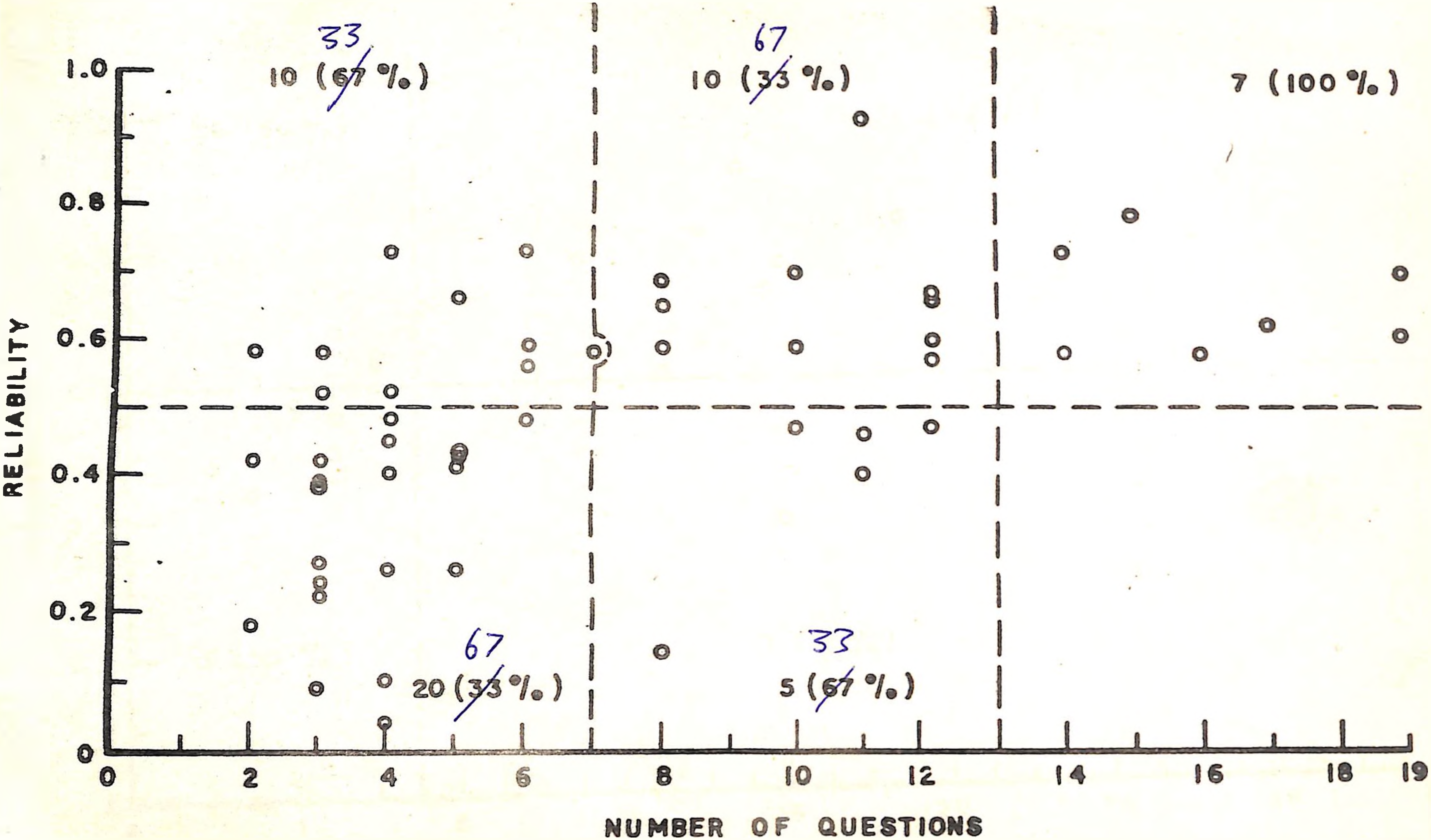


FIG. 3.3a DEMARCATION OF THRESHOLD VALUES FOR THE NUMBER OF QUESTIONS AS ALSO THE RELIABILITY LEVELS FOR THE TESTS.

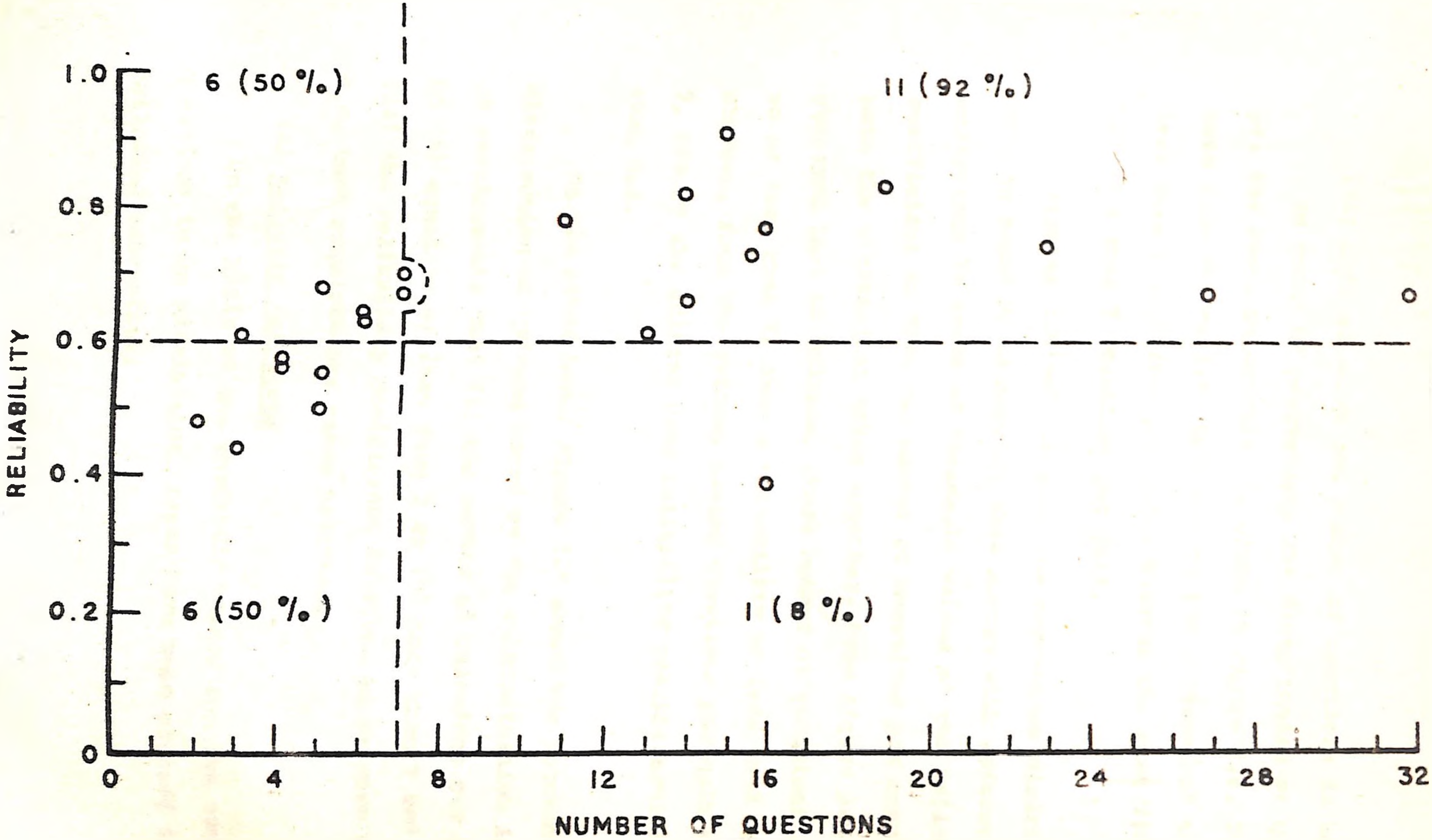


FIG. 3.3b DEMARCATION OF THRESHOLD VALUES FOR THE NUMBER OF QUESTIONS AS ALSO THE RELIABILITY LEVEL FOR THE COMPREHENSIVES.

(ii) Quizzes where the number of questions is more than 7.

In terms of percentage, the distribution of quizzes, as per the above categories, is given in figure 3.4. It can be seen from figure 3.4, that 29% of the quizzes had either 7 or less than 7 questions per quiz, whereas the rest 71% of quizzes had more than 7 questions per quiz.

Figures 3.5 and 3.6 give the percentage distribution of quizzes based on categories that emerge with reference to the constraints in terms of threshold values of the reliability coefficient as also the number of questions per examination, both the constraints taken together. From figure 3.5, it is observed that no quizzes, where number of questions is equal to or less than 7, show a reliability of less than 0.4, whereas, from the quizzes having questions per quiz more than 7, 20% of the quizzes have reliability coefficients less than 0.4.

On the other hand, figure 3.6 shows the percentage distribution of quizzes based on the categorization in terms of requirements that (i) the number of questions per quiz be (a) equal to or less than 7 or (b) more than 7 and that (ii) the reliability coefficient be equal to or more than 0.4, both requirements taken together.

(c) Analysis for Tests

On the basis of the threshold values for the number of questions in an examination, tests have been grouped into following categories:

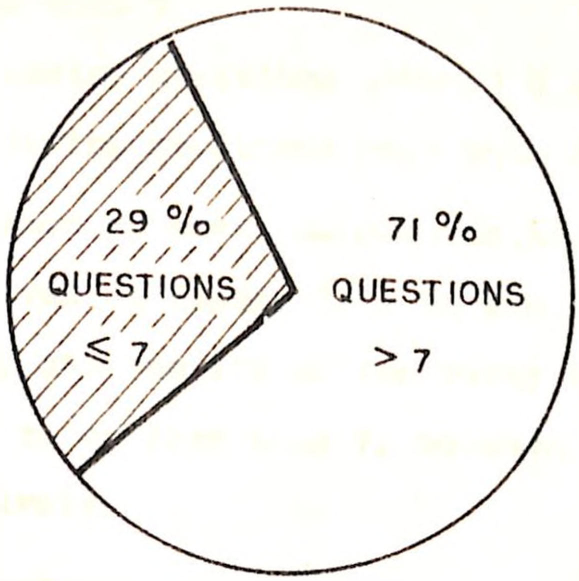


FIG. 3·4 DISTRIBUTION OF QUIZZES ACCORDING TO NUMBER OF QUESTIONS.

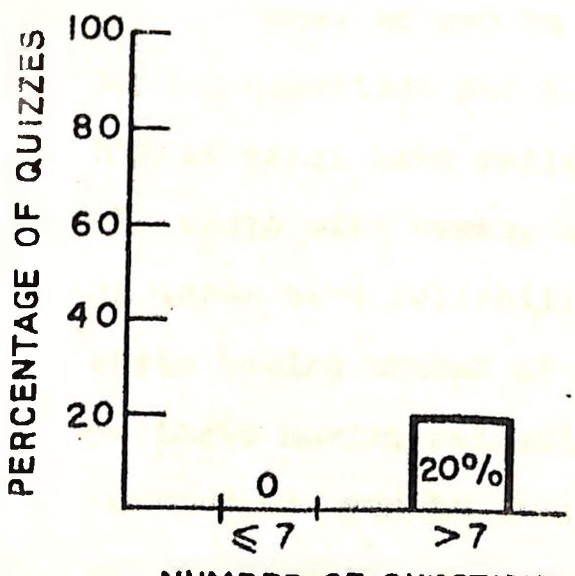


FIG. 3·5 PERCENTAGE OF QUIZZES WITH R < 0.4.

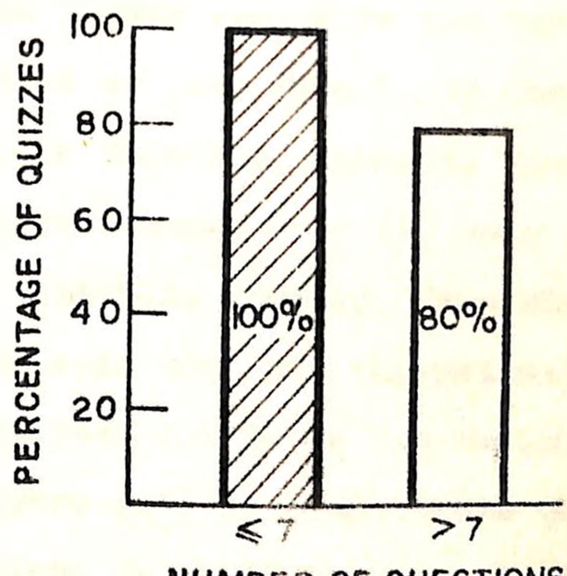


FIG. 3·6 PERCENTAGE OF QUIZZES WITH R ≥ 0.4.

(i) Tests where the number of questions is equal to or less than 7

(ii) Tests having questions between 8 and 13

(iii) Tests having questions more than 13.

Distribution of tests as per the aforementioned categories is given in figure 3.7. As can be seen from figure 3.7, 58%, 29% and 13% of the tests had number of questions equal to or less than 7, between 8 and 13 and more than 13, respectively.

Figures 3.8 and 3.9 show the percentage distribution of tests based on the various categories that emerge with reference to the constraints in terms of threshold values of (i) the reliability coefficient as also (ii) the number of questions in a test, both types of constraints taken together.

Thus, as can be seen from figure 3.8, from the tests having questions per test equal to or less than 7, as many as 57% of tests have reliability less than 0.5, whereas, from the tests with number of questions between 8 to 13, only 33% of tests have reliability less than 0.5, and, finally, from the tests having number of questions more than 13, the percentage of tests having reliability less than 0.5 is as low as 0%. Further, as can be seen from figure 3.9, from the tests where per examination number of questions is equal to or less than 7, only 33% of tests have reliability equal to or more than 0.5, whereas, from the tests having number of questions between 8 to 13, 67% (tests) have reliability equal to or more than 0.5, and , finally, from the tests having number of questions

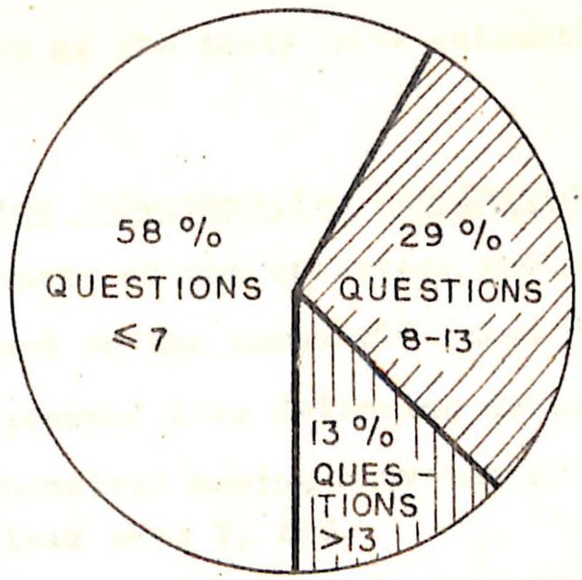


FIG. 3·7 DISTRIBUTION OF TESTS ACCORDING TO NUMBER OF QUESTIONS.

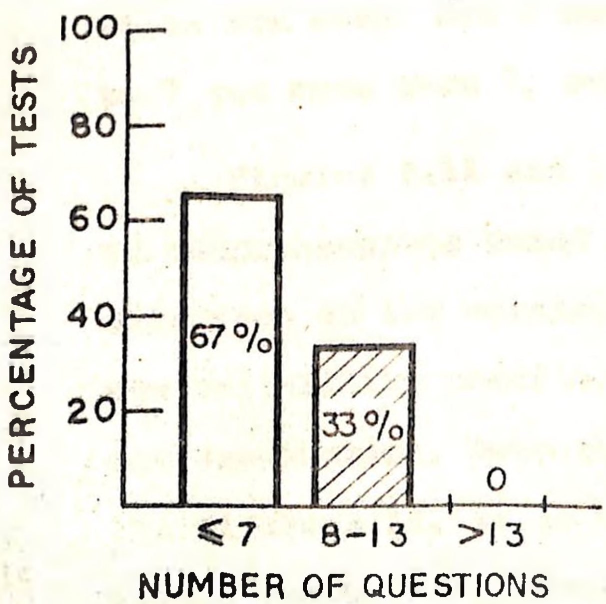


FIG. 3·8 PERCENTAGE OF TESTS WITH $R < 0.5$.

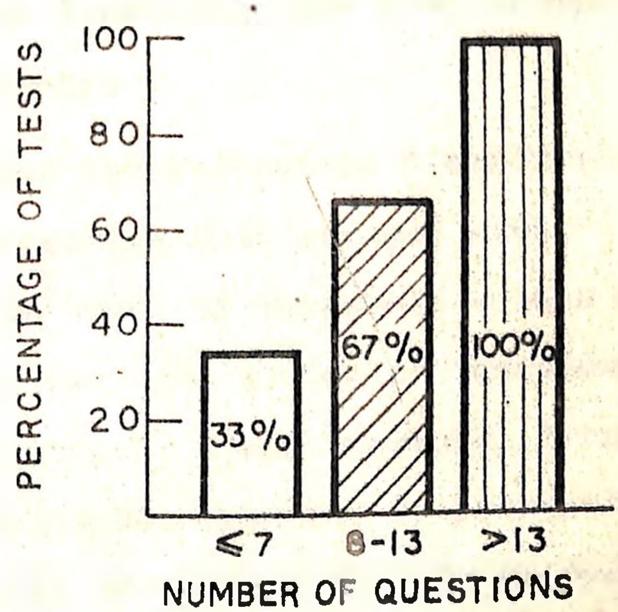


FIG. 3·9 PERCENTAGE OF TESTS WITH $R \geq 0.5$.

more than 13, 100% of the tests have reliability equal to or more than 0.5.

(d) Analysis for Comprehensive Examinations

For the purpose of the analysis, the comprehensive examinations, based on the threshold value for the number of questions, were grouped into following categories:

(i) Comprehensives having a number of questions equal to or less than 7, and

(ii) Comprehensives having a number of questions more than 7.

The percentage distribution of comprehensives according to the above categorization is shown in Figure 3.10. The figure shows, that, an exactly equal percentages of comprehensives (i.e. 50% ~~each~~) had a number of questions less than or equal to 7 and more than 7, ~~respectively~~.

Figures 3.11 and 3.12 give the percentage distribution of comprehensives based on categories that emerged with reference to the constraints in terms of threshold values of the reliability coefficient as also the number of questions per examination, both the constraints taken together. From the figure 3.11, it is observed that, from the comprehensives having number of questions equal to or less than 7, as many as 50% have reliability less than 0.6, while, from the comprehensives having number of question more than 7, only 8% have reliability less than 0.6.

Against this, from the figure 3.12, it can be seen that, for the comprehensives having less than or equal to 7 questions,

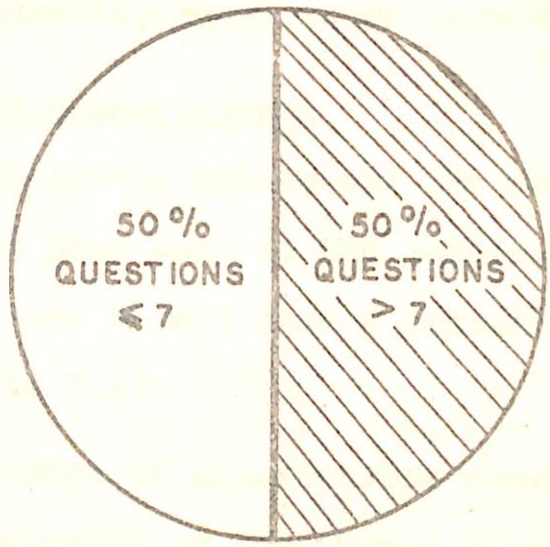


FIG. 3·10 DISTRIBUTION OF COMPREHENSIVES ACCORDING TO NUMBER OF QUESTIONS.

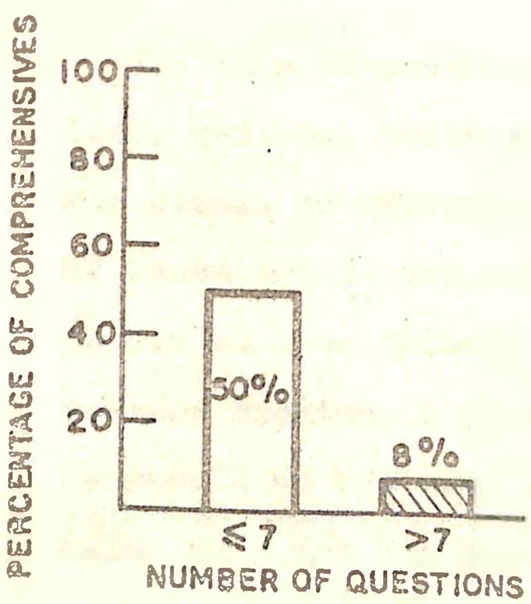


FIG. 3·11 PERCENTAGE OF COMPREHENSIVES WITH R < 0.6.

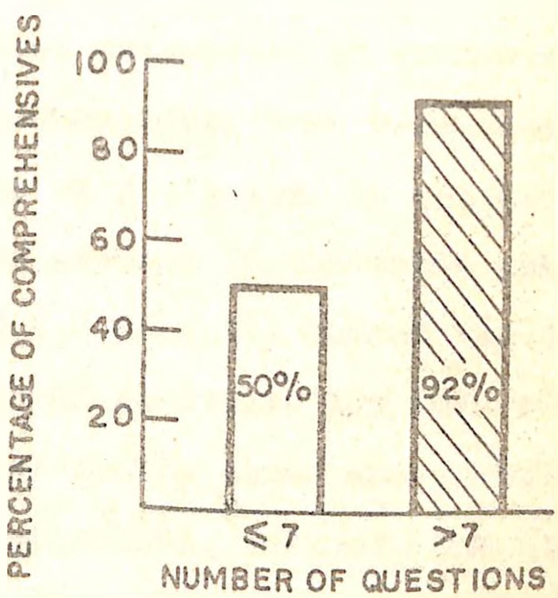


FIG. 3·12 PERCENTAGE OF COMPREHENSIVES WITH R ≥ 0.6.

50% have reliability equal to or more than 0.6, whereas, from the comprehensives having more than 7 questions, as many as 92% have reliability equal to or more than 0.6.

(e) Summary of observations:

(1) At a macro-level, taking all the examinations together, it has been found that most of the examinations where number of questions is more than 7 show a satisfactory reliability (0.5, see figure 3.1 to 3.3).

(2) The percentage of examinations showing a reliability of less than 0.5 decreases as the number of questions increases (see figure 3.2). As a natural corollary to this, from figure 3.3, it emerges that the percentage of examinations having reliability equal to or more than 0.5 increases as the number of questions increases.

(3) At a micro-level, various categories of examinations, i.e., quizzes, tests and comprehensives, have been considered. The sample of quizzes consists of 7 quizzes, as against the 52 tests and 24 comprehensives studied in the above context. As can be seen through tables 3.4 to 3.22, out of the seven quizzes studied, 6 have number of questions per quiz ranging between 8 to 12, and, thus, invariably above the threshold value of 7 for the number of questions. Further, out of the remaining two, one has the number of questions equal to 7 and the remaining quiz, also, has its number of questions, though lower than the threshold value, more or less equal to it, the same being 6. Thus, from these details it ~~seems to~~ emerges that, on an average, there is a tendency (on the part of teachers) to give fairly

good number of questions in the quizzes. Indeed, this observation is also corroborated from figure 3.19, from where one can see that the average of the number of questions per quiz lies between the averages for the same for tests and quizzes.

Further, from figures 3.5 and 3.6, one can see that, whatever may be the number of questions-per-quiz-wise categorization of the quizzes, for 80% or above quizzes the reliability is equal to or more than 0.4. Thus, the set of quizzes studied has, on the whole, demonstrated a satisfactory reliability level.

When the above two observations in terms (i) a tendency to give a good number of questions in the quiz and (ii) the satisfactory reliability level as demonstrated by the quizzes, are put together, it then, ~~seems to~~ emerge~~s~~, as in the case of the analysis for the all examinations taken together, that higher the number of questions in an examination, greater is the probability of its having a higher reliability.

In the context of the above observation, one may be tempted to further critically study figures 3.5 and 3.6 to see how the percentage of quizzes for $R < 0.4$ or $R \geq 0.4$ varies with increasing number of questions in a quiz. But, then, it turns out that the sample size at hand to pursue any such further analysis is indeed very small, in the sense that in an already small sized sample of 7 quizzes, only one quiz has number of questions less than the threshold value of 7 and, yet, there, too, the number of questions, as already indicated, is as high as 6. Thus, it can be seen that the data at hand is just not adequate to draw any further detailed observations.

(4) Coming to the analysis for tests, from figure 3.8, it can be seen that the percentage of tests having low reliability (i.e., less than 0.5) decreases as the number of questions increases. Vice-versa, from figure 3.9, it clearly emerges that the percentage of tests having reliability equal to or more than 0.5 increases as the number of questions per test increases.

(5) It is also clear that, like the tests, the percentage of comprehensives showing a reliability of less than 0.6 decreases as the number of questions increases (figure 3.11). Against this, from figure 3.12, it can be seen that the percentage of comprehensives with reliability equal to or greater than 0.6 increases as the number of questions in comprehensives increases.

(6) Thus, in summary, at the macro-level in terms of 'all the examinations taken together' as also at the micro-level in terms of the 'tests' and 'comprehensives', it clearly emerges that the percentage of examinations with reliability equal to or more than satisfactory increases as the number of questions per examination increases, thus implying that higher the number of questions, higher the reliability. As for the quizzes, though they (quizzes) could not be analysed explicitely in terms of the relationship between the increase in the number of questions per quiz and the corresponding

increase in the degree of reliability, they have, also, demonstrated clearly that higher the number of questions per examination, greater the possibility of the examination having the satisfactory reliability.

Thus, it can be concluded that higher the number of questions in an examination, greater is the probability of the examination having the higher reliability.

In addition to the earlier observations it may be noted that the number of questions and thus the range of questions shows an increasing order from the quizzes to comprehensives as is obvious from following table:

Table 3.24

Range of Questions in Various Examinations

Examinations	Range of Questions
Quizzes	6-12
Tests	2-19
Comprehensives	2-32

(ii) Reliability Observations According to the Duration of Examinations

(a) The Analysis

The duration of an examination is one of the important factors that goes to describe different categories of examinations such as quizzes, tests and comprehensives. The distribution of examinations according to duration is shown in figure 3.13, from where it can be seen that the total set of examinations selected for the study consisted of 8% quizzes (having duration of about 10-20 minutes), 63% tests(having duration of about 1 hour), the rest 29% being comprehensives (having duration of about 2-3 hours). The percentages of quizzes, tests and comprehensives having reliability of less than 0.5 are given in figure 3.14, while figure 3.15 gives the percentages for above categories of examinations having reliability equal to or more than 0.5. Thus, while, from figure 3.14, one can see that for reliability less than 0.5, the percentage of quizzes (57%) is higher than that for the tests (48%), is higher than that for comprehensives (12%), from the figure 3.15, it clearly emerges that, for reliability equal to or greater than 0.5, the corresponding percentages of quizzes, tests and comprehensives vary from 43% to 52% to 88%, respectively.

Figure 3.16 gives the average reliability plotted against the duration of examinations. Thus, from the figure 3.16, it can be seen that as the examination time duration varies from 10-20 minutes to 1 hour to 2-3 hours, the average reliability coefficient goes from 0.55 to 0.5 to 0.6, respectively.

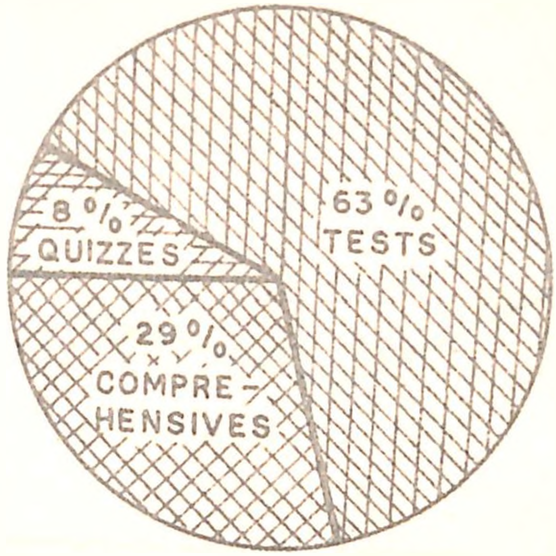


FIG. 3.13 DISTRIBUTION OF EXAMINATIONS ACCORDING TO QUIZZES, TESTS AND COMPREHENSIVES.

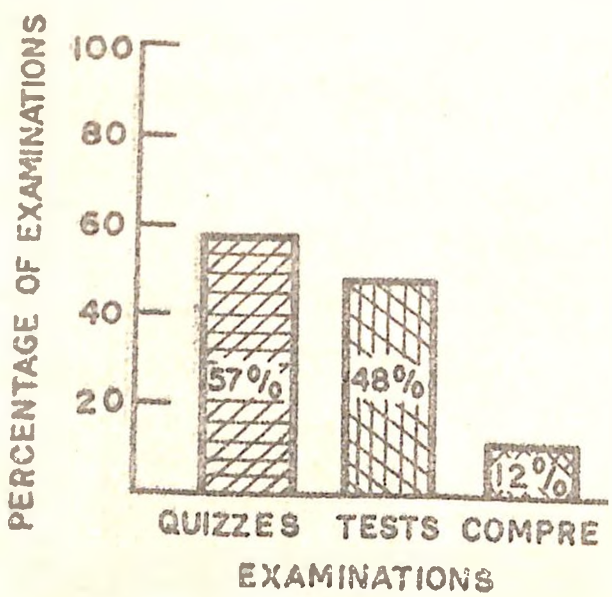


FIG. 3.14 PERCENTAGE OF EXAMINATIONS WITH $R < 0.5$.

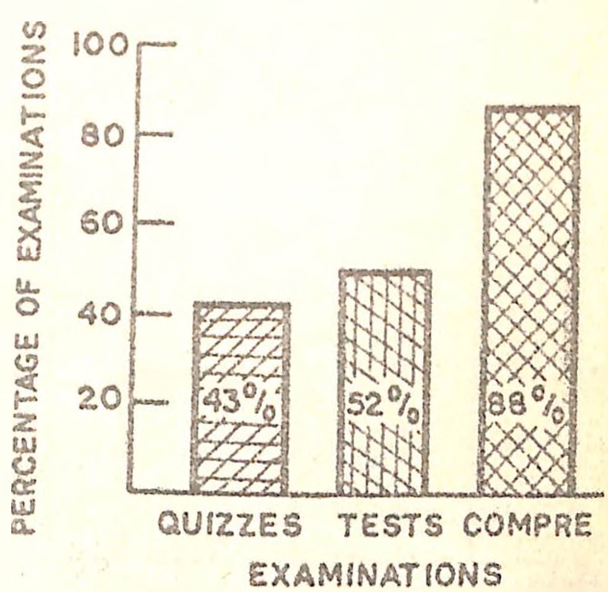


FIG. 3.15 PERCENTAGE OF EXAMINATIONS WITH $R \geq 0.5$.

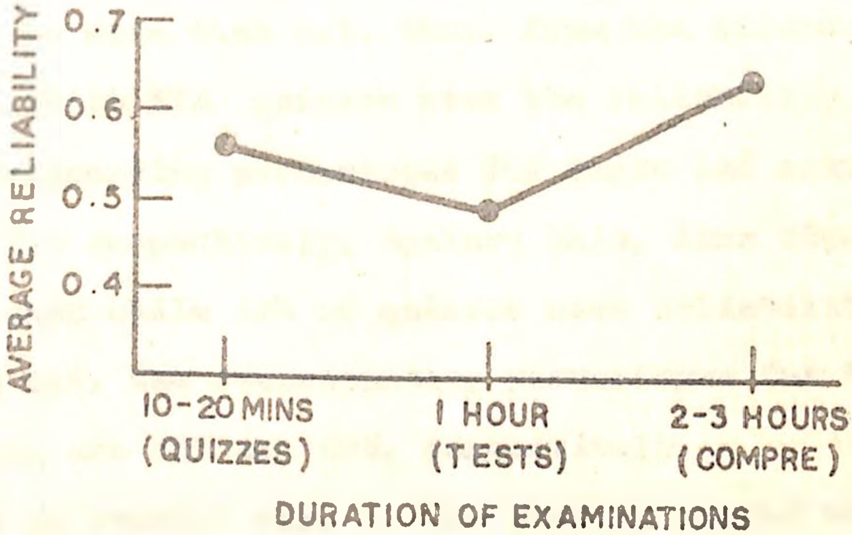


FIG. 3.16 AVERAGE RELIABILITY PLOTTED AGAINST DURATION OF EXAMINATIONS.

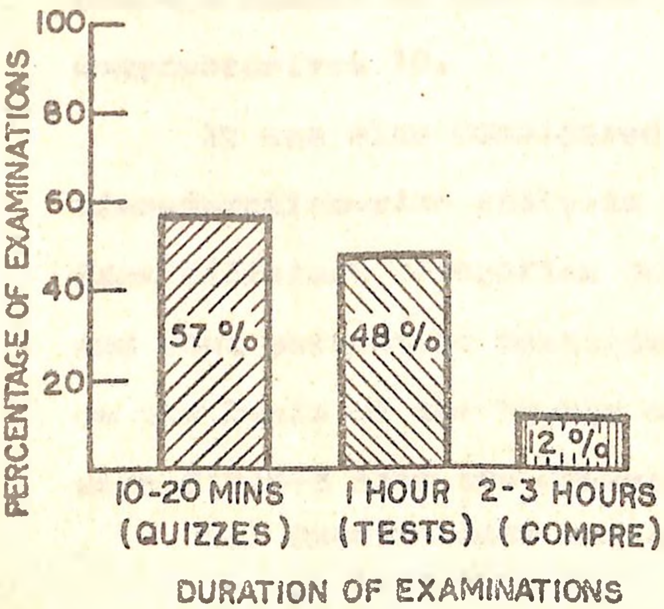


FIG. 3.17 PERCENTAGE OF EXAMINATIONS WITH $R < 0.5$.

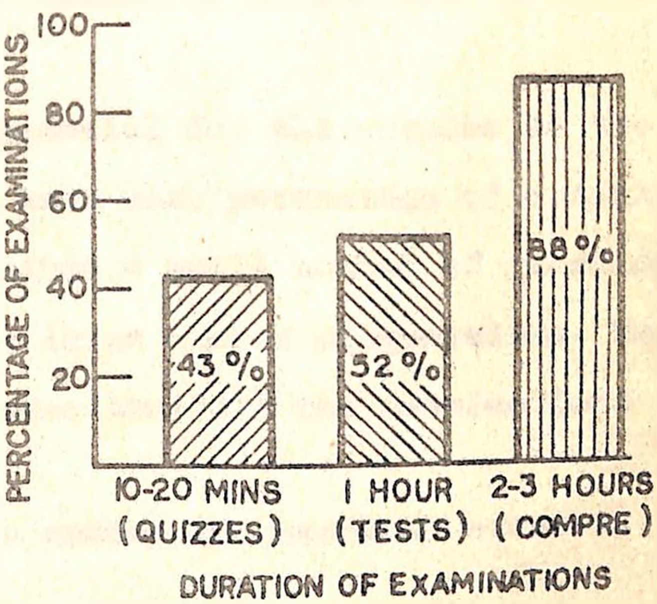


FIG. 3.18 PERCENTAGE OF EXAMINATIONS WITH $R \geq 0.5$.

For each of the above mentioned categories of examinations, figure 3.17* describes what percentage examinations have reliability less than 0.5, while figure 3.18* gives percentage examinations under each category when the reliability coefficients are equal to or more than 0.5. Thus, from the figure 3.17, one can see that while 57% quizzes have the reliability less than 0.5, the corresponding percentages for tests and comprehensives are 48% and 12% respectively. Against this, from figure 3.18, one can see that while 43% of quizzes have reliability equal to or more than 0.5, the corresponding percentages for tests and comprehensives are 52% and 88%, respectively.

To get an overall view of the average number of questions in different examinations, a graph has been plotted showing an average number of questions against the duration of examinations (see figure 3.19). Thus, as can be seen from this figure, the average number of questions in quizzes is 9, in tests 8 and in comprehensives 10.

It was also considered essential for the purpose of the time-duration-wise analysis to know what percentage of examinations from different categories contained a small number of questions and what percentage contained a large number of questions. Hence, on the basis of the number of questions all the examinations were divided into two groups.

(i) Examinations having a number of questions equal to or less than 7,

(ii) Examinations having a number of questions more than 7.

---

* Figure 3.17 is same as figure 3.14 and figure 3.18 is same as figure 3.15. This repetition has taken place because of the one to one correspondence between the practice followed at the Institute in terms of the traditional nomenclature for the examinations and the time durationwise classification for the same.

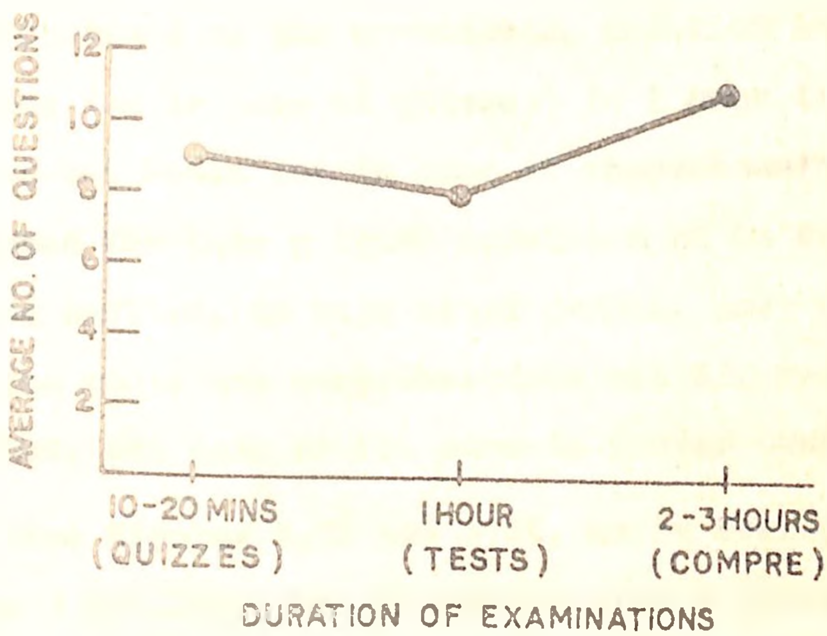


FIG. 3·19 AVERAGE NUMBER OF QUESTIONS PLOTTED AGAINST DURATION OF EXAMINATIONS.

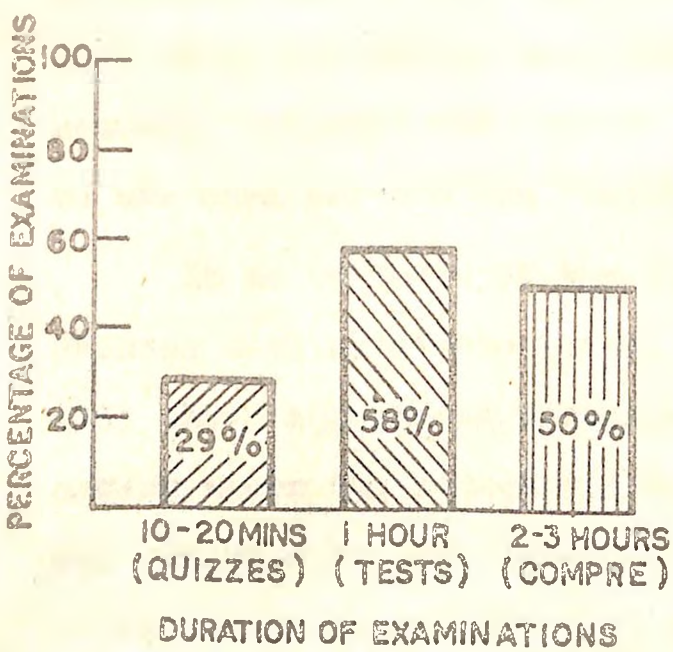


FIG. 3·20 PERCENTAGE OF EXAMINATIONS WITH AVERAGE NUMBER OF QUESTIONS ≤ 7.

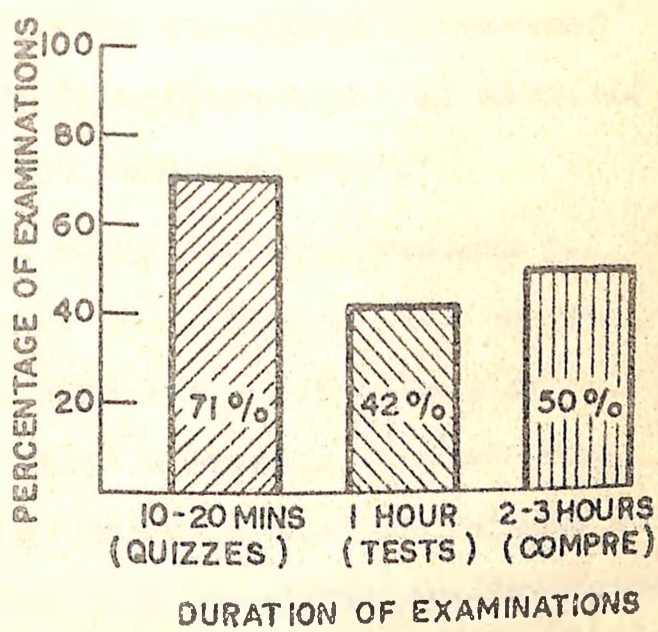


FIG. 3·21 PERCENTAGE OF EXAMINATIONS WITH AVERAGE NUMBER OF QUESTIONS > 7.

8 This can be seen from figure 3.20 and 3.21, respectively.

(b) Summary of Observations

1. Figure 3.16 gives the variation in the averages of reliability coefficients as the examination duration increases from 10-20 minutes (as in case of quizzes) to 1 hour (as in case of tests) to 2-3 hours (as in case of comprehensives). The sample size studied for this purpose consisted of 83 examinations, and, as discussed earlier, in this total sample, only 8% are quizzes, while the tests and comprehensives are 63% and 29%, respectively. Thus, the quiz sample size is indeed very small.

Further, from figures 3.20 and 3.21, on an average, one can also observe a tendency for quizzes to have a fairly good number of questions per quiz. This could be because usually teachers opt to give objective and short answer type questions in quizzes (see figure 2.5). A natural outcome of all this is that there may always be a substantial structural difference between 'quizzes' and 'tests' and 'comprehensives' as compared to the same between the 'tests' and 'comprehensives'.

It is in terms of the above mentioned observations on quizzes with reference to (i) their small sample size as also (ii) their structural difference from the examinations as coming under the categories of 'tests & comprehensives' that one can then see the inadequacy of the data-base on quizzes in terms of the same being used along with the data-base for tests and comprehensives to study the contents of figure 3.16.

Against the above perspective then, limiting one's attention to the data-base on tests and comprehensives only, one can then observe from figure 3.16 that the average reliabilit increases from 0.5 to 0.6 as the time duration increases from the order of 1 hour to that of 2-3 hours.

2. The above observation on reliability vis-a-vis duration of examination is further confirmed when one studies figure 3.18, from where one can see that the percentages of examinations having reliability more than or equal to 0.5 increases from 43% to 52% to 88% as the examination duration changes from 10-20 minutes to 1 hour to 2-3 hours, respectively. Figure 3.17 gives an alternate statement in the above context, as it shows that percentage of examinations having reliability less than 0.5 decreases from 57% to 48% to 12%, as the duration increases from 10-20 minutes to 1 hour to 2-3 hours, respectively.

3. Thus, the observation emerges that higher the duration of the examination, greater the possibility of the reliability being higher.

4. It may be mentioned that, implicitly, this observation further goes to strengthen the observation in terms of the relationship between the reliability of an examination and the number of questions in an examination, as higher the examination duration, greater would be the probability of the number of questions in the examination being higher.

Indeed, accounting for the tendency to give a fairly good number of questions in the quizzes, from figure 3.19, one can easily see in terms of tests and comprehensives the validity of the above proposition pertaining to the relationship between the time duration of an examination and the number of questions in it.

(iii) Reliability Observations According to the Number of Students

For the purpose of the analysis, different types of examinations i.e. the evaluation components such as quizzes, tests and comprehensives have been grouped into different categories based on the threshold values for the student number in an examination (for illustration see figures 3.21a and 3.21b). Thus, for the examinations coming under the type of 'tests and comprehensives', these values workout to be in all two in number, namely, 100 and 150. For the quizzes, the threshold value emerging in terms of the number of students in an examination is 50.

As a result, for tests and comprehensives, the different groups emerging as per the threshold values on the student number are:

(i) Examinations where the number of students is less than 100.

(ii) Examinations where the students are between 100-150.

(iii) Examinations where the number of students is more than 150.

For quizzes, the categories are as follows:

(i) Quizzes where the number of students is less than 50.

(ii) Quizzes where the number of students is equal to or more than 50.

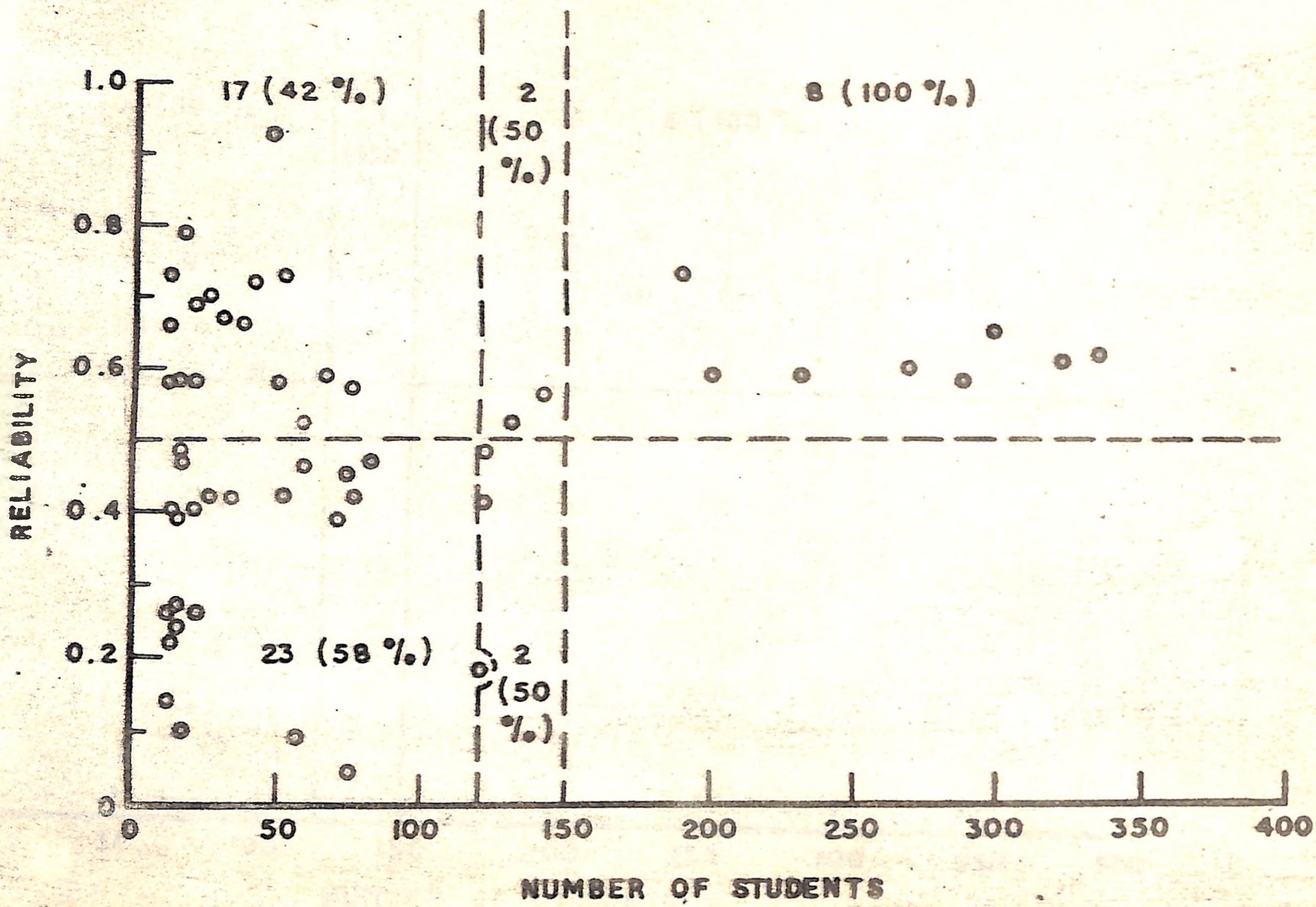
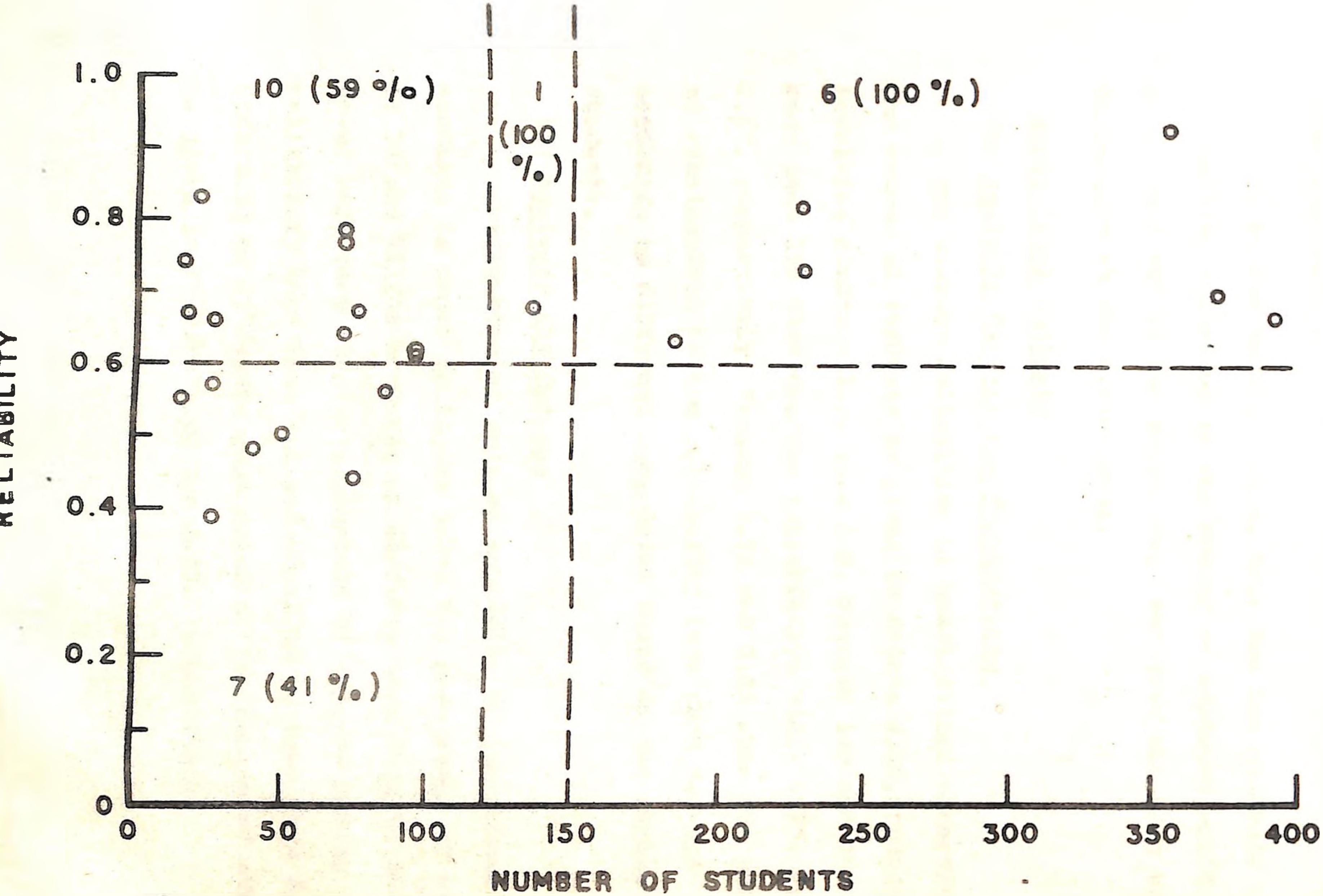


FIG. 3.21a DEMARCATION OF THRESHOLD VALUES FOR THE NUMBER OF STUDENTS AS ALSO THE RELIABILITY LEVEL FOR THE TESTS.



. 3·21 b DEMARCATION OF THRESHOLD VALUES FOR THE NUMBER OF STUDENTS AS ALSO THE RELIABILITY LEVEL FOR THE COMPREHENSIVES.

Categories for tests and comprehensives as also for the quizzes, according to the threshold values for reliability, are ~~the~~ same as the ones mentioned in the beginning of this sub-section .

As in the earlier cases, here too the analysis of reliability according to the number of students would first be carried out at the macro-level and then the same would be pursued at the micro-level.

Macro-level Analysis

(a) Analysis for All the Examinations

The average reliability of examinations according to the number of students is given in figure 3.22. Examinations involving students less than 100, between 100 and 150 and more than 150 show average reliability, viz., 0.52, 0.46 and 0.67, respectively. Figures 3.23 and 3.24 show the percentages of examinations having reliability less than 0.5 and $R \geqslant 0.5$ according to different categories based on the number of students.

(b) Analysis for Quizzes

Distribution of quizzes according to the number of students is shown in figure 3.25. The percentage of examinations is 29% and 71% for the number of students less than or equal to 50 and ~~equal to or~~ more than 50 respectively. The percentage of quizzes having a reliability less than 0.4 and equal to or more than 0.4, according to different categories of the number of students, is given in figures 3.26 and 3.27, respectively.

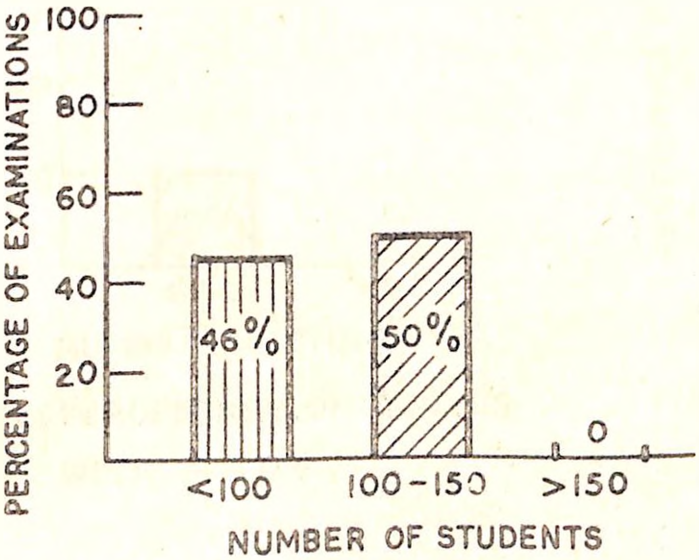


FIG. 3.23 PERCENTAGE OF EXAMINATIONS WITH $R < 0.5$.

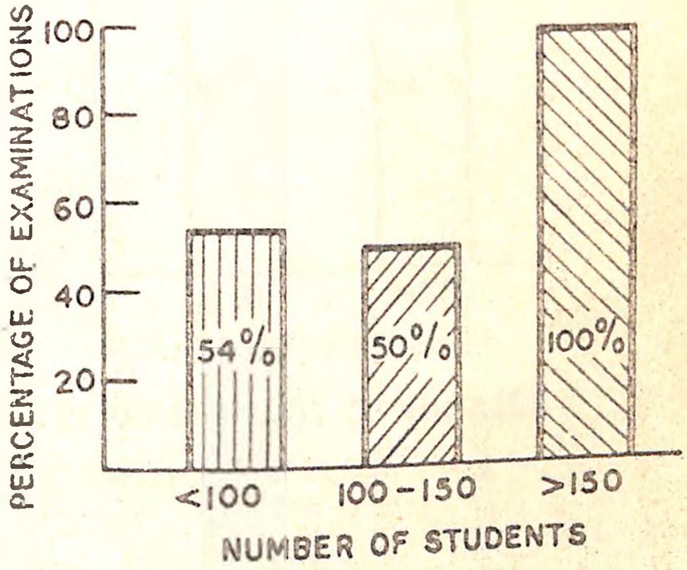


FIG. 3.24 PERCENTAGE OF EXAMINATIONS WITH $R \geqslant 0.5$.

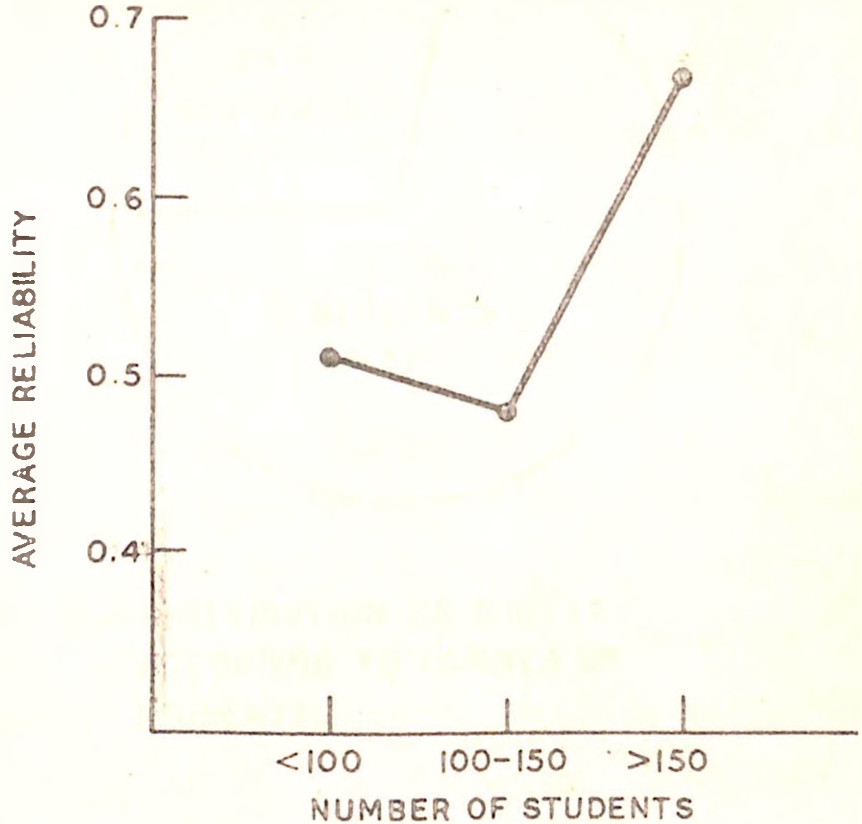


FIG. 3.22 AVERAGE RELIABILITY PLOTTED AGAINST NUMBER OF STUDENTS.

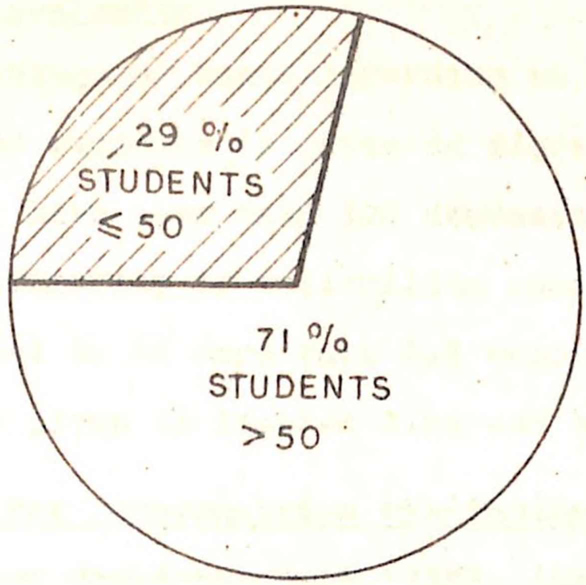


FIG. 3·25 DISTRIBUTION OF QUIZZES ACCORDING TO NUMBER OF STUDENTS.

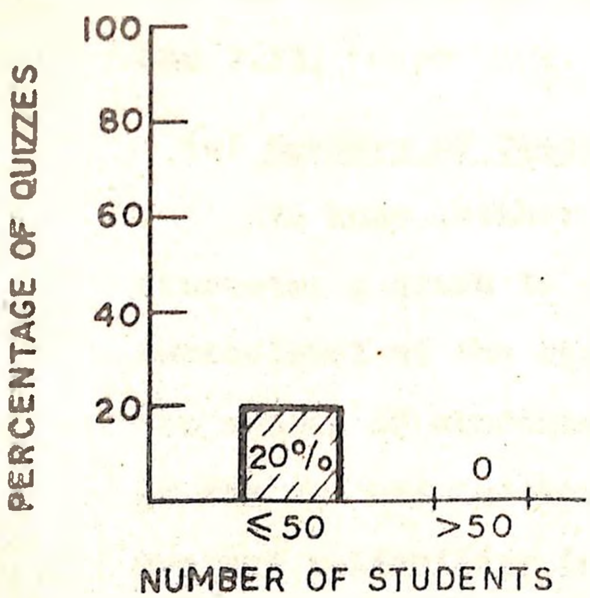


FIG 3·26 PERCENTAGE OF QUIZZES WITH $R < 0.4$.

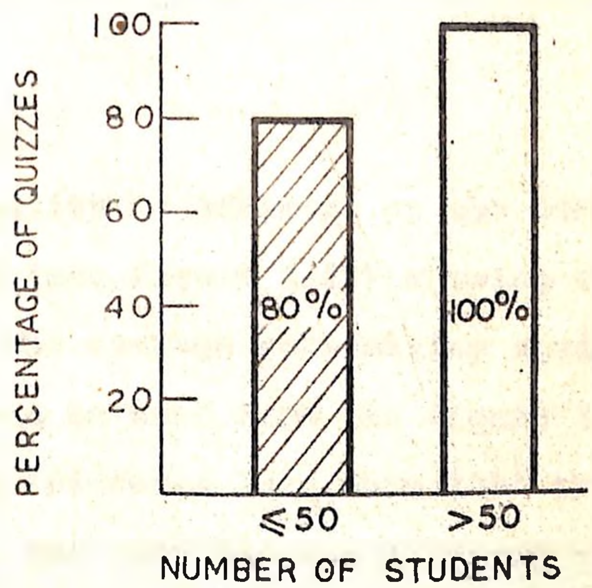


FIG 3·27 PERCENTAGE OF QUIZZES WITH $R \geq 0.4$.

(c) Analysis for Tests

The percentage of tests according to different categories of the number of students is given in figure 3.28. Majority of the tests (77%) have less than 100 students. Distributions of tests in the categories of reliability less than 0.5 and reliability equal to or more than 0.5 according to the number of students are given in figures 3.29 and 3.30, respectively.

(d) Analysis for Comprehensive examinations

Most of the comprehensives (71%), like the tests, have a number of students less than 100. The distribution of comprehensives according to the number of students can be seen from figure 3.31. Percentage-wise distribution of comprehensives according to number of students having reliability less than 0.6 and equal to or more than 0.6 are given in figures 3.32 and 3.33, respectively.

(e) Summary of Observations

To know whether reliability is affected by the number of students, a graph is plotted (see figure 3.22) showing at macro-level of the analysis the average reliability against the number of students. As can be seen from the figure 3.22, if for the examinations with 'students less than 100' the average reliability is 0.51, the same for examinations with 'students between 100-150' is 0.48 and in case of examinations with 'students more than 150', the average reliability works out to be 0.68.

Coming to some further details, 77% of the total sample studied in the above context had 'student number less than

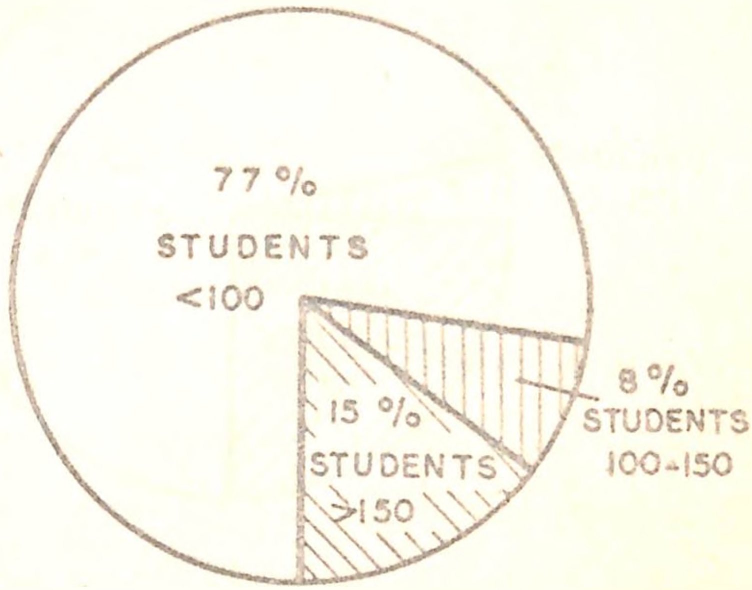


FIG. 3.28 DISTRIBUTION OF TESTS ACCORDING TO NUMBER OF STUDENTS.

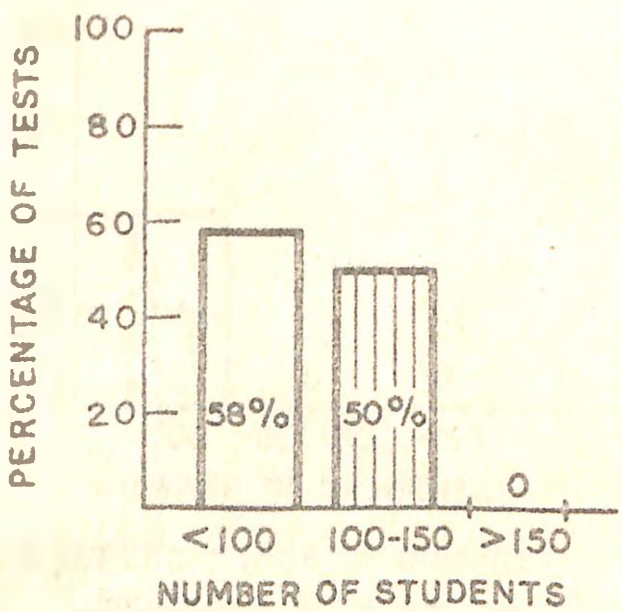


FIG. 3.29 PERCENTAGE OF TESTS WITH R < 0.5.

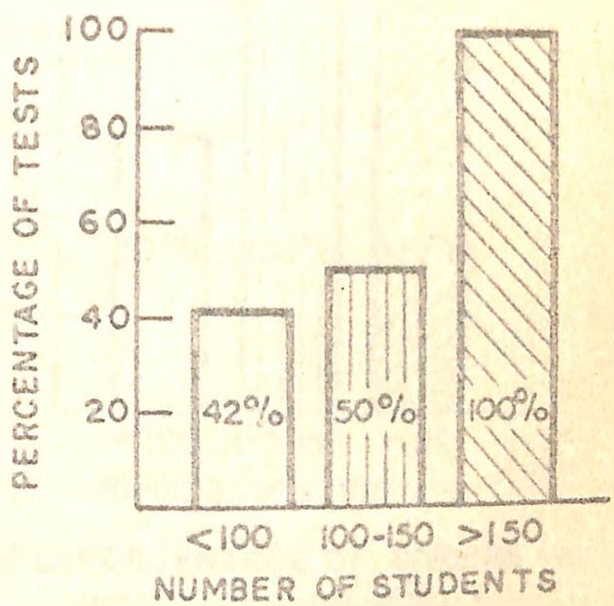


FIG. 3.30 PERCENTAGE OF TESTS WITH R ≥ 0.5.

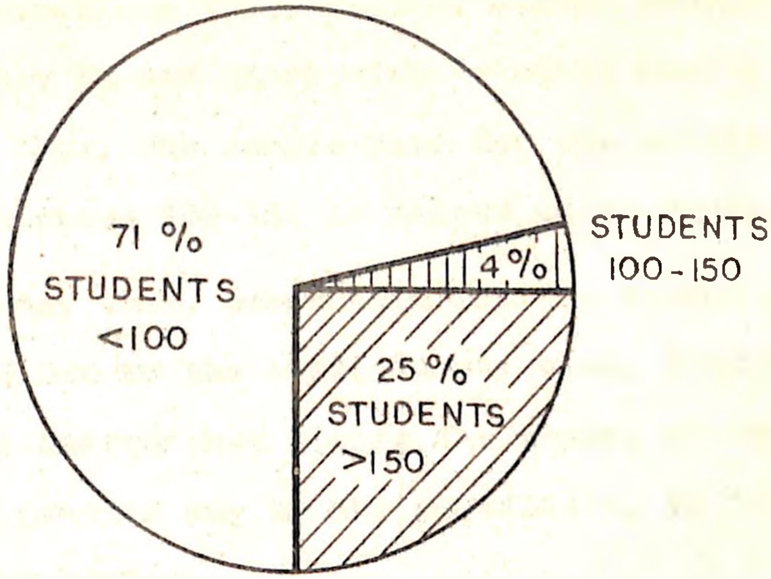


FIG. 3·31 DISTRIBUTION OF COMPREHENSIVES ACCORDING TO NUMBER OF STUDENTS.

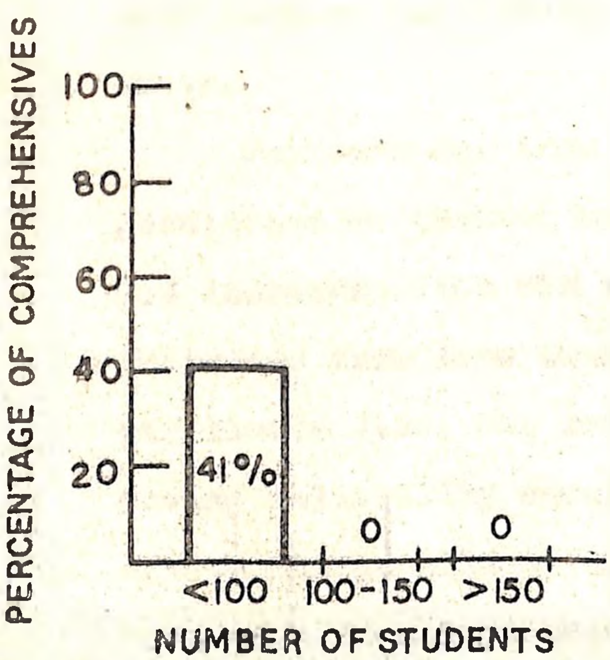


FIG. 3·32 PERCENTAGE OF COMPREHENSIVES WITH R < 0.6.

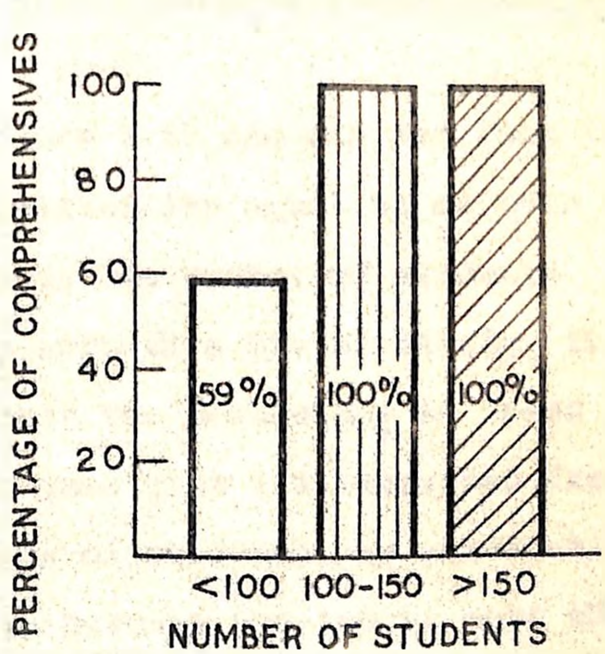


FIG. 3·33 PERCENTAGE OF COMPREHENSIVES WITH R ≥ 0.6.

100', while examinations 'with student number between 100-150' accounted for only 6% and those with 'student number more than 150' were 17%. Thus, the sample size for the examinations with student number between 100-150 is indeed quite small.

In such case, then, accounting for the random aberration that is possible due to the small sample size, from figure 3.22 it ~~seems to~~ emerges that higher the number of students in an examination, greater may be the possibility of its reliability being higher.

Figures 3.23 and 3.24 also in an overall context convey the same observation as above.

At a micro level, quizzes, tests and comprehensives, when studied separately, also give a similar observation as above.

For example, from the figure 3.27, one can see that the percentage of quizzes having reliability equal to or more than 0.4 increases from 80% to 100% as the number of students increases from less than or equal to 50 to more than 50. Similarly, from the figure 3.30, one can see that the percentage of tests having reliability equal to or more than 0.5 increases from 42% to 50% to 100% as the number of students per examination increases from less than 100 to between 100-150 to more than 150. And, finally, from figure 3.33 it clearly emerges that the percentage of comprehensives having reliability equal to or more than 0.6 increases from 59% to 100% to 100%, as the number of students increases from less than 100 to between 100-150 to more than 150.

Thus it ~~seems to~~ emerges that higher the number of students in the examination, greater could be the probability of the examination having a satisfactory reliability.

A further critical analysis in the above context reveals interesting observations. To begin with, it seems to be that normally the examinations for courses attended by a large number of students are also characterized by a large number of questions. This could be because the courses attended by a large number of students are usually at the core level and it is at the core level that teachers seem to have tendency to give large number of questions in an examination (see figure 3.46). Thus, the higher reliability for examinations characterized by larger number of students can then be explained in terms of the observation that they (the examinations) invariably seem to have higher number of questions. This, in turn, also then goes to further strengthen the proposition in terms of the relationship between the reliability of an examination and the number of questions in it.

(iv) Reliability Observations According to the Level of Courses

(a) Analysis for All the Examinations Taken together

Figure 3.34 shows the variation in the average reliability of examinations when plotted against the various course levels. Thus, for the sample size of 83 examinations studied, it works out that as one moves from the I level to II to III to IV to V level courses, the average reliability varies from 0.62 to 0.92 to 0.52 to 0.48 to 0.5, respectively. Thus, on the whole,

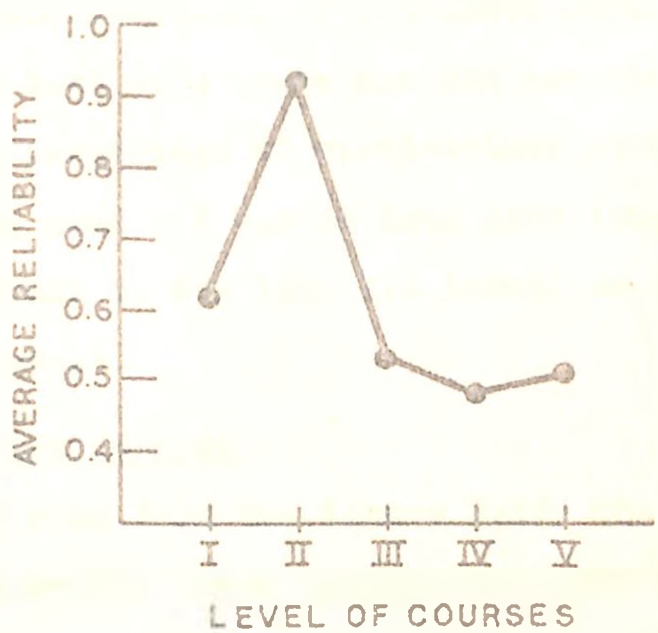


FIG. 3·34 AVERAGE RELIABILITY OF EXAMINATIONS PLOTTED AGAINST LEVEL OF COURSES.

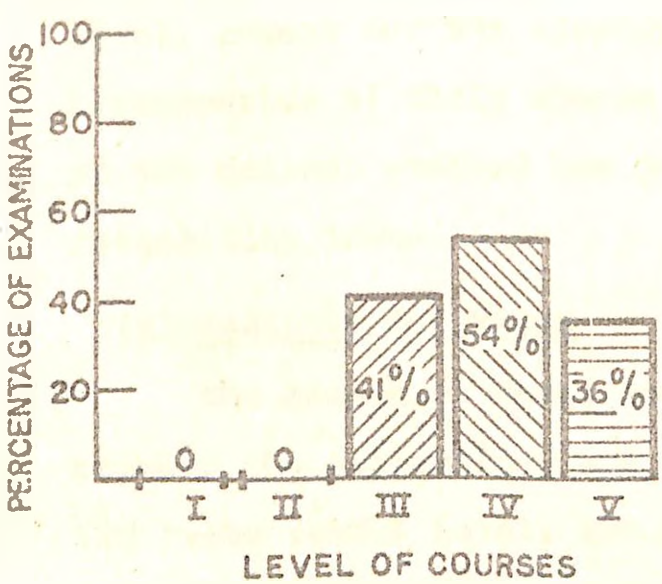


FIG. 3·35 PERCENTAGE OF EXAMINATIONS WITH $R < 0.5$.

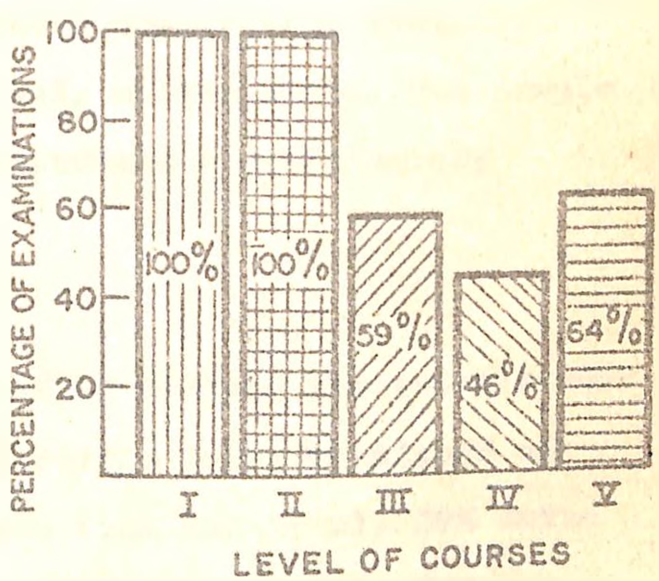


FIG. 3·36 PERCENTAGE OF EXAMINATIONS WITH $R \geq 0.5$.

the average reliability seems to decrease as one moves from the foundation or core level courses to the higher or professional level courses.

An alternate statement of the above observation emerges from the figure 3.36 from where one can see that from I level to V level, the percentage of examinations having reliability equal to or more than 0.5 varies from 100% (for I level) to 100% (for II level) to 59% (for III level) to 46% (for IV level) to 64% (for V level).

(b) Analysis for Quizzes

As can be seen from the figure 3.37, the sample of quizzes studied in the above context consists of 43% quizzes each from the III and IV levels and 14% quizzes from the V level. Indeed, when one studies figures 3.38 and 3.39, nothing particular seems to emerge from the sample of quizzes in terms of the relationship between the reliability and the course level, except for the already known observation that, irrespective of their course level, on the whole, the sample of the quizzes studied has demonstrated a satisfactory reliability level.

(c) Analysis for Tests

The sample of tests studied to analyse the relationship between the reliability and the course level consisted of 12% tests from I level, 23% tests from III level, 29% tests from the IV level and 36% tests from the V level. For the sample so distributed, from figure 3.42 one can see that as one moves from I level to the V level, the percentage of

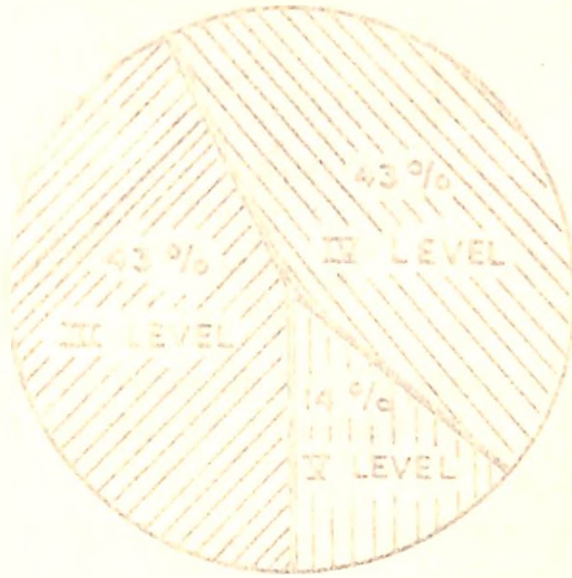


FIG. 3·37 DISTRIBUTION OF QUIZZES ACCORDING TO LEVEL OF COURSES.

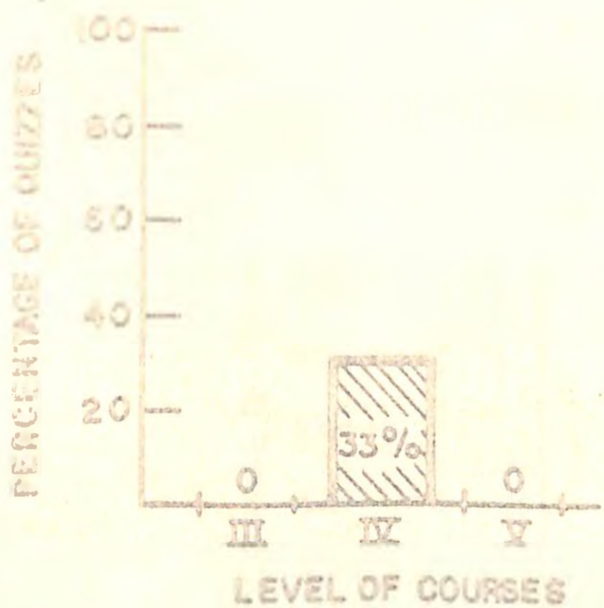


FIG. 3·38 PERCENTAGE OF QUIZZES WITH $R < 0.4$.

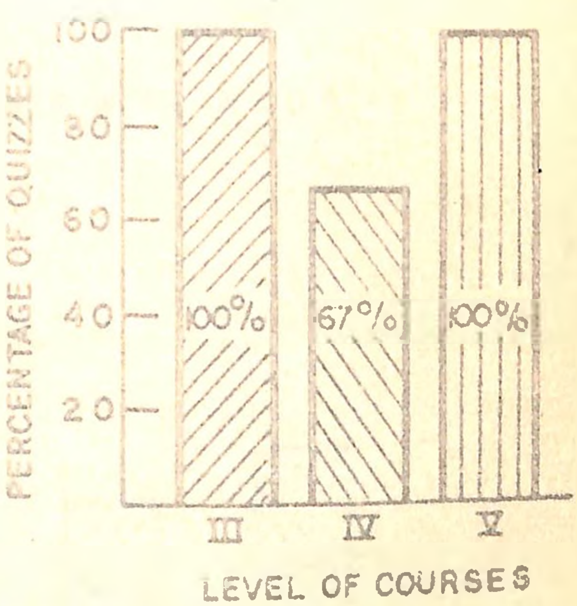


FIG. 3·39 PERCENTAGE OF QUIZZES WITH $R \geq 0.4$.

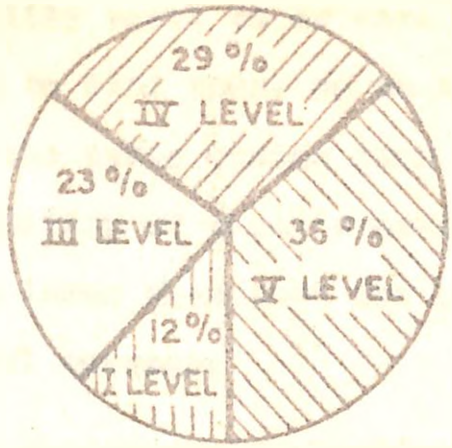


FIG. 3.40 DISTRIBUTION OF TESTS ACCORDING TO LEVEL OF COURSES.

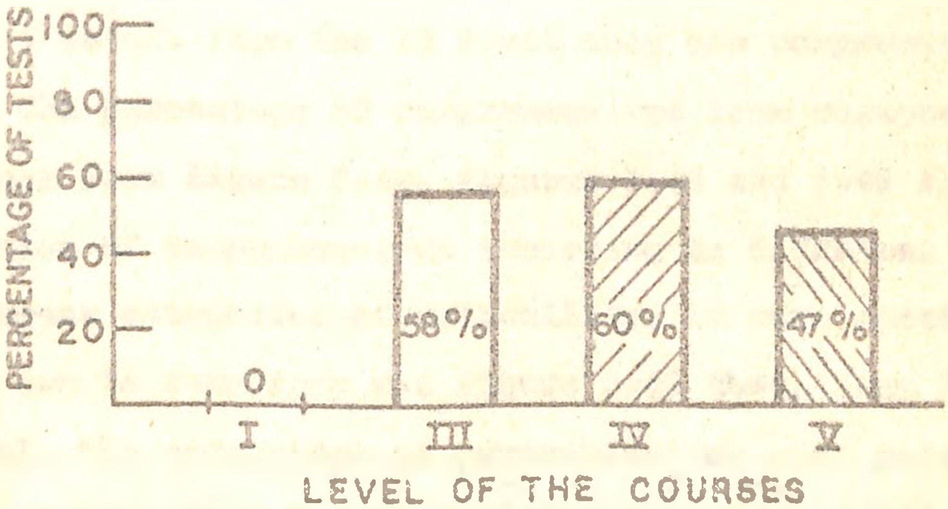


FIG. 3.41 PERCENTAGE OF TESTS WITH $R < 0.5$.

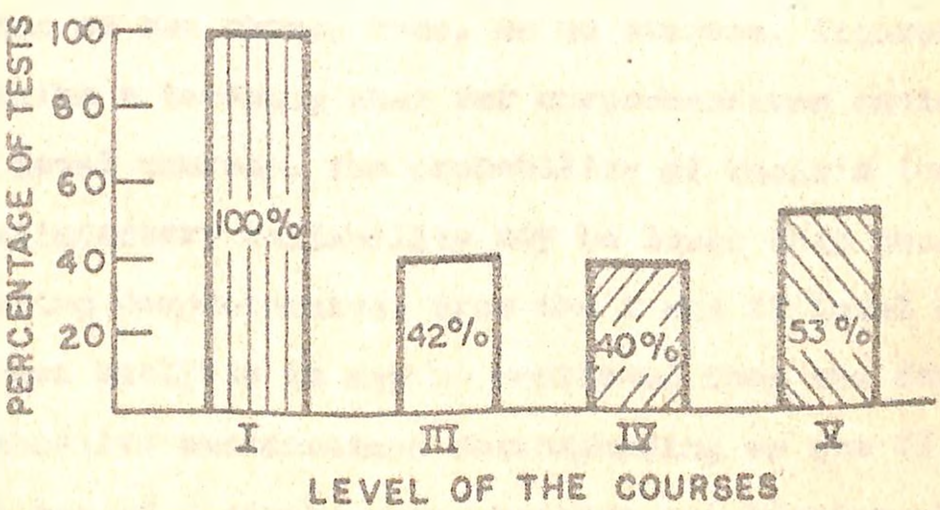


FIG. 3.42 PERCENTAGE OF TESTS WITH $R \geq 0.5$.

tests with reliability equal to or more than 0.5 varies from 100% to 42% to 40% to 53%; thus, on an average, describing a tendency that for the tests coming from III or IV or V level courses, the probability of their's having a satisfactory reliability may be lower than that for the corresponding tests from say the I level courses.

(d) Analysis for Comprehensive examinations

The comprehensives selected for study represent all the levels of the course. The maximum number of comprehensives were from the V level. From the II level only one comprehensive was studied. The percentage of comprehensives from various levels can be seen from figure 3.44. Figures 3.44 and 3.45 show the distribution of comprehensives according to different levels and different categories of reliability. In more specific terms, then, it can be seen from the figure 3.45 that, from I level to V level, the percentage of comprehensives with reliability equal to or more than 0.6 varies from 100% (for I level) to 100% (for II level) to 71%(for III level) to 83% (for IV level) to 50%(for V level). Figure 3.44 provides a complimentary observation of the above. Thus, on an average, figures 3.45 and 3.46 describe a tendency that for comprehensives coming from III to V level courses, the probability of their's (comprehensives having satisfactory reliability may be lower than that for the corresponding comprehensives from the I and II level courses. In the above analysis it may be mentioned that the data in terms of comprehensive examinations corresponding to the II level is only in terms of a single comprehensive examination. This was mainly due to the non availability of larger data. Even though there is

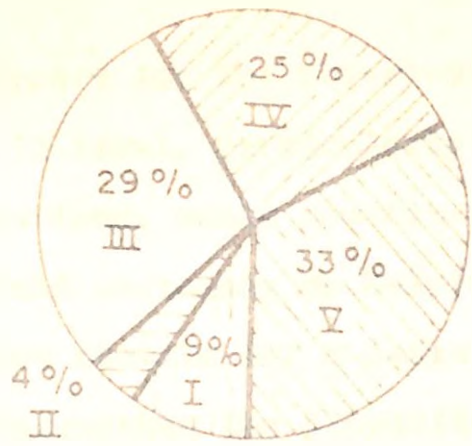


FIG 3·43 DISTRIBUTION OF COMPREHENSIVES ACCORDING TO LEVEL OF COURSES

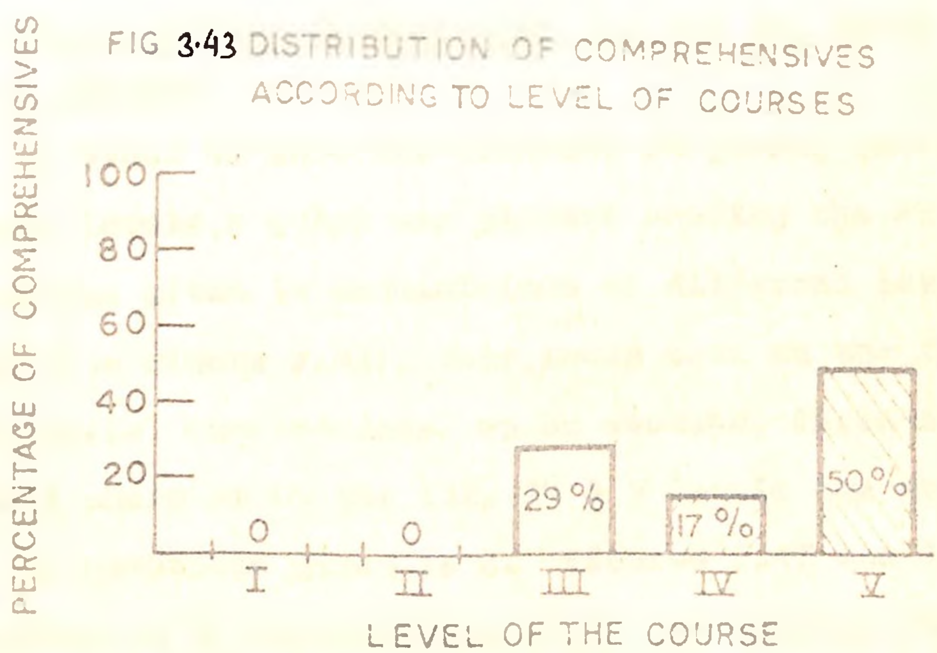


FIG. 3·44 PERCENTAGE OF COMPREHENSIVES WITH R<0.6

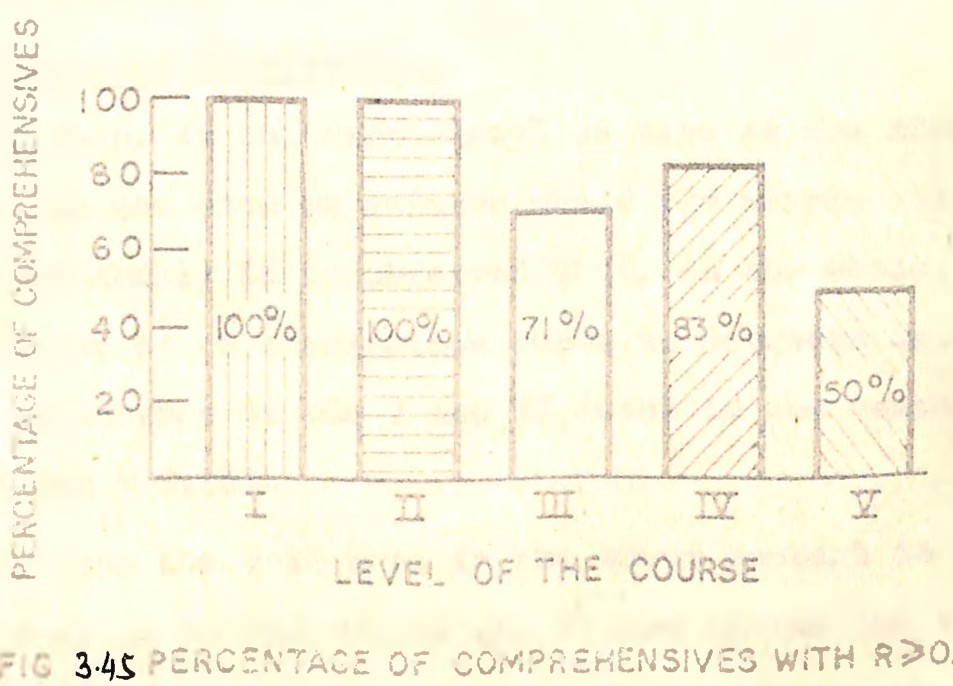


FIG 3·45 PERCENTAGE OF COMPREHENSIVES WITH R≥0.6.

emerging a wide support for the observation that normally comprehensives at II level, particularly due to their having a good number of questions, would demonstrate a sabisfactory reliability, it would certainly be helpful in the futuristic sense to study these details for a larger number of examinations so as to further strengthen the proposition under consideration.

(e) <u>Average Number of Questions Against the Level of Courses</u>

In order to know the tendency on giving questions at different levels, a graph was plotted showing the average number of questions given in examinations at different levels of courses (see Figure 3.46). This graph shows that at the first and second levels examinations, on an average, included 15 questions where as at the III, IV & V levels the average number of questions given is 8. Figures 3.47 and 3.48 show the percentage of examinations having an average number of questions equal to or less than 7 and more than 7, respectively across the various levels of the courses from I to V.

(f) <u>Summary of Observations</u>

1. Thus, at the macro-level as also at the micro-level (except in the case of quizzes where the sample size is found to be too small) it is observed that, on the whole, the average reliability of an examination seems to decrease as one moves from say courses at the I and II level to the courses at say the IV and V level.

2. When the data base in the above context is further pursued so as to understand the reason behind the type of

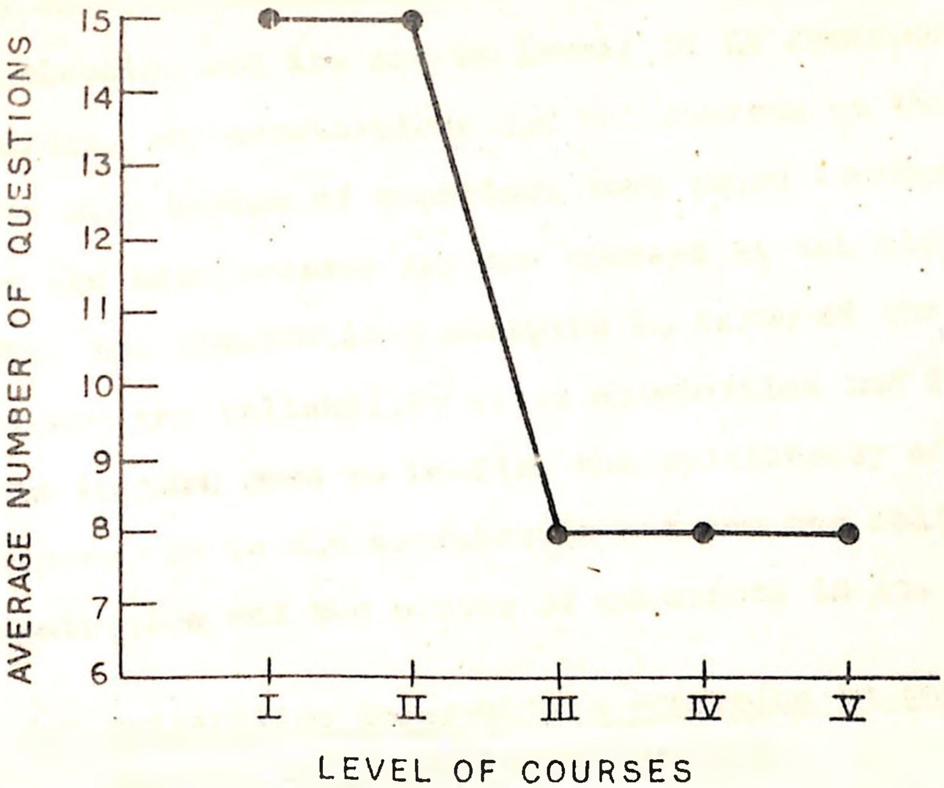


FIG.3·46 AVERAGE NUMBER OF QUESTIONS IN EXAMINATIONS PLOTTED AGAINST LEVEL OF COURSES.

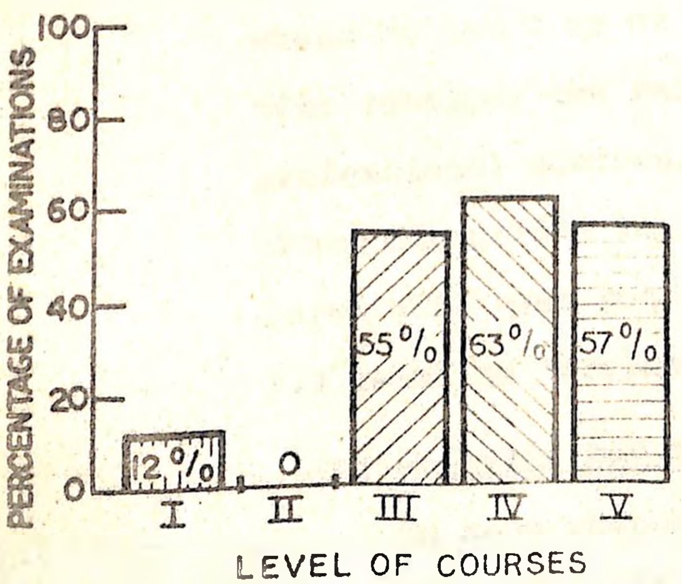


FIG.3·47 PERCENTAGE OF EXAMINATIONS WITH AVERAGE NUMBER OF QUESTIONS ≤ 7.

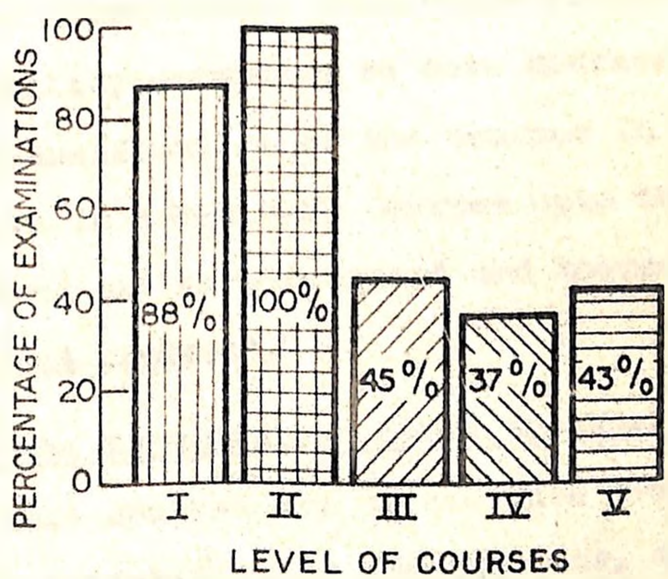


FIG.3·48 PERCENTAGE OF EXAMINATIONS WITH AVERAGE NUMBER OF QUESTIONS > 7.

the emerging correlation between the reliability of an examination and its course level, it is observed that, on an average, the examinations for the courses at the lower level have more number of questions than those (number of questions) for the examinations ~~for~~ in the courses at the higher level. Thus, the observation~~s~~ emerging in terms of the correlation between the reliability of an examination and its course level also in turn goes to confirm the consistency of the earlier observation on the correlation between the reliability of an examination and the number of questions in it.

(v) Reliability Observations According to the Core Courses and Professional Courses

At the micro-level, observations on reliability of an examination against its course level have been discussed ~~in the~~ above ~~preceding section.~~ Thus here ~~In the present sub-section~~ the analysis would be taken up at the macro-level. Specifically one can then ~~Thus, this sub-section will~~ consider the reliability according to core courses and professional courses. Classification of the courses in the above categories is given in table 3.2. Courses upto the III level have been considered as 'Core Courses' and beyond the III level as 'Professional Courses'.

(a) Analysis for All the Examinations taken together

To know whether core courses are better than professional courses in terms of the reliability of examinations, a graph has been plotted showing average reliability of examinations against the level of courses (see figure 3.49). For the core and professional levels, the percentages of examinations having

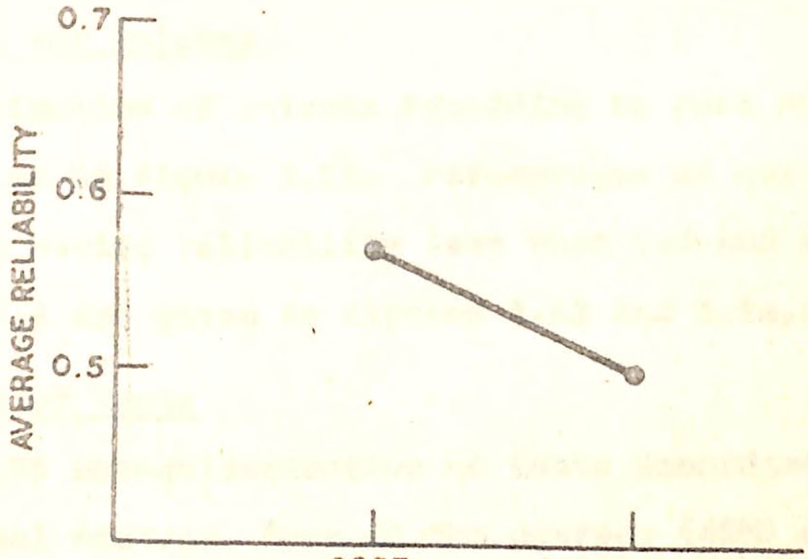


FIG. 3.49 AVERAGE RELIABILITY OF EXAMINATIONS PLOTTED AGAINST LEVEL OF COURSES.

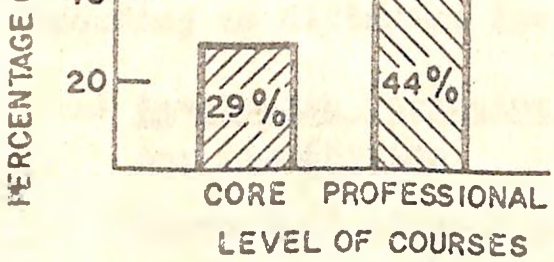


FIG. 3.50 PERCENTAGE OF EXAMINATIONS WITH $R < 0.5$.

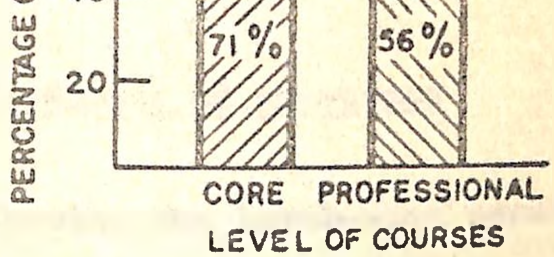


FIG. 3.51 PERCENTAGE OF EXAMINATIONS WITH $R \geq 0.5$.

reliability less than 0.5 and reliability equal to or more than 0.5 are given in figure 3.50 and 3.51, respectively.

(b) Analysis for Quizzes

The distribution of quizzes according to core and professional courses is given in figure 3.52. Percentages of quizzes from various levels having reliability less than 0.4 and equal to or more than 0.4 are given in figures 3.53 and 3.54, respectively.

(c) Analysis for Tests

Figure 3.55 shows distribution of tests according to core and professional courses. Most of the courses (65%) were from the professional level. Percentage of tests from core and professional courses having reliability lessthan 0.5 and equal to or more than 0.5 are given in figures 3.56 and 3.57, respectively.

(d) Analysis for Comprehensive Examinations

The percentage of comprehensives according to core and professional courses is shown in figures 3.58. Figures 3.59 and 3.60 indicate the percentages of comprehensives having a reliability of less than 0.6 and equal to or morethan 0.6, according to different levels.

(e) Level-wise Variation in the Number of Questions per Examination

Figure 3.61 gives a graph showing the level-wise variation in the average number of questions per examination. Thus, ~~as~~ while average number of questions in an examination given in core courses is 10, ~~where as~~ it is 8 in case of professional courses. Figures 3.62 and 3.63, for the core and professional levels, give percentages of

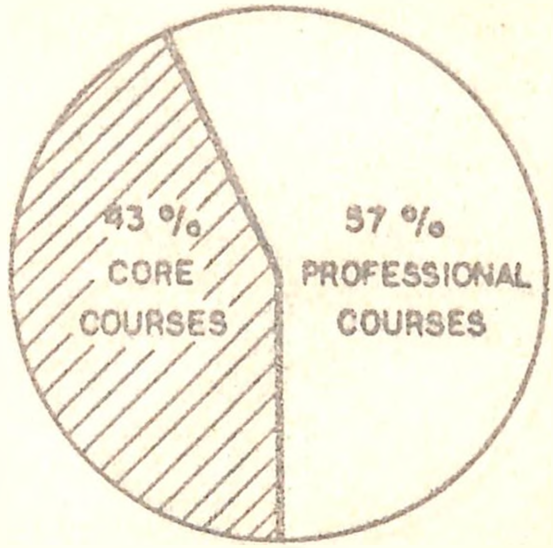


FIG. 3.52 DISTRIBUTION OF QUIZZES ACCORDING TO LEVEL OF COURSES.

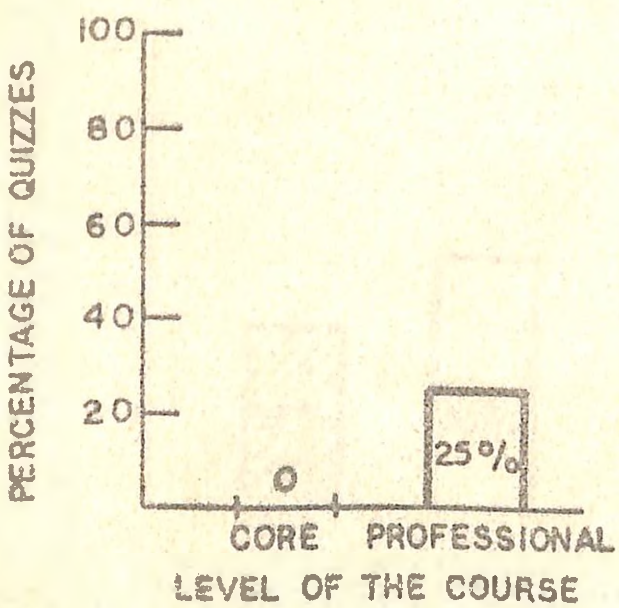


FIG. 3.53 PERCENTAGE OF QUIZZES WITH R<0.4.

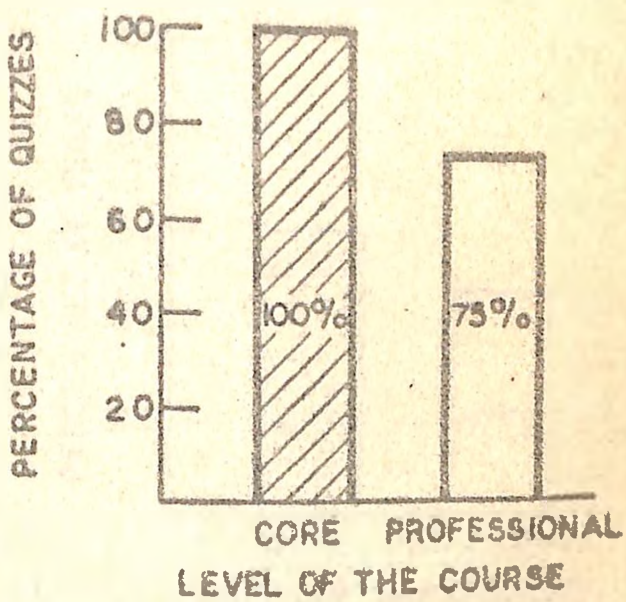


FIG. 3.54 PERCENTAGE OF QUIZZES WITH R ≥ 0.4.

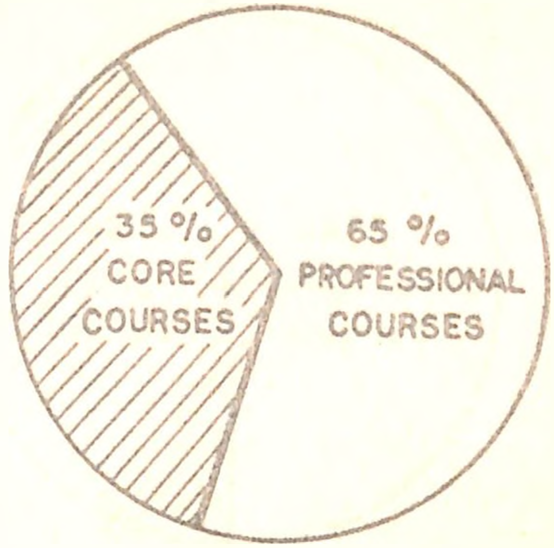


FIG. 3.55 DISTRIBUTION OF TESTS ACCORDING TO LEVEL OF COURSES.

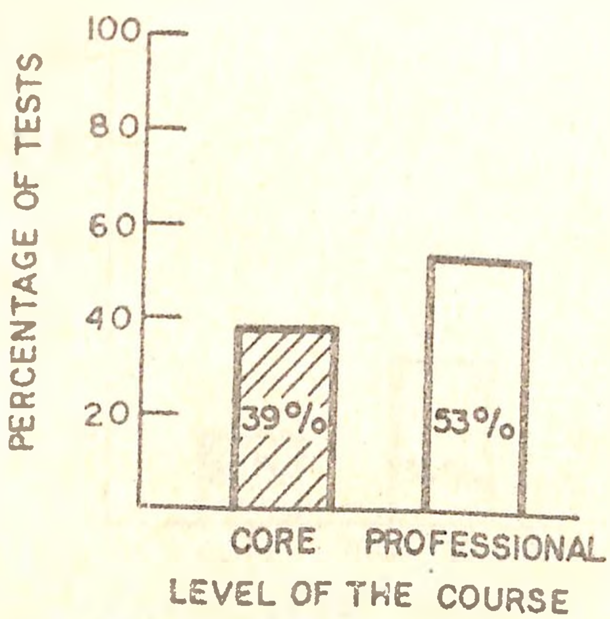


FIG. 3.56 PERCENTAGE OF TESTS WITH $R < 0.5$.

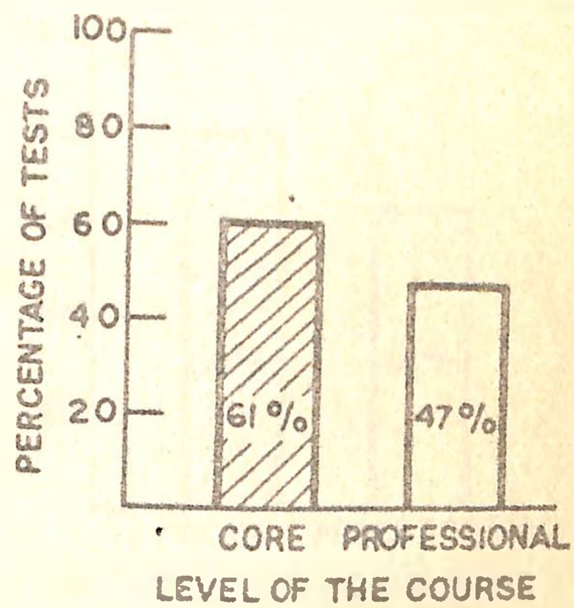


FIG. 3.57 PERCENTAGE OF TESTS WITH $R \geqslant 0.5$.

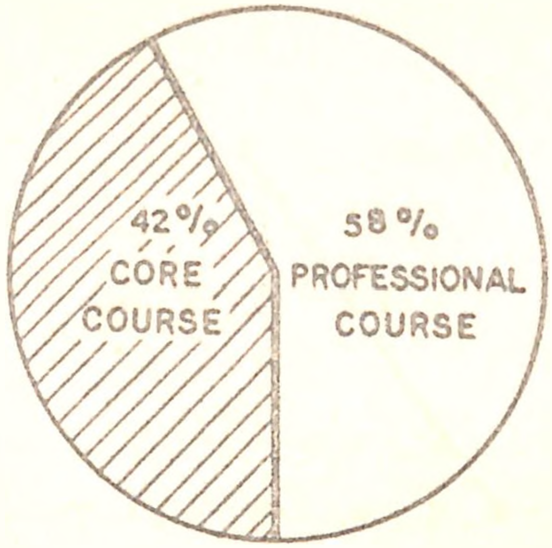


FIG. 3.58 DISTRIBUTION OF COMPREHENSIVES ACCORDING TO LEVEL OF COURSES.

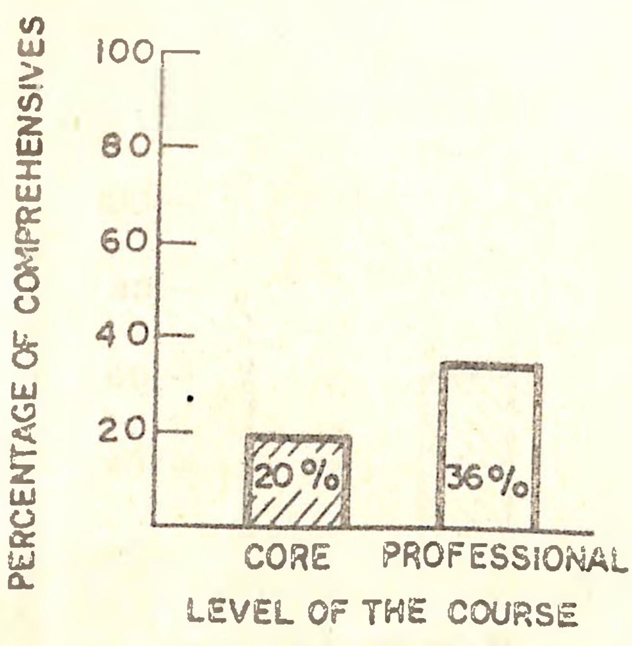


FIG. 3.59 PERCENTAGE OF COMPREHENSIVES WITH R < 0.6.

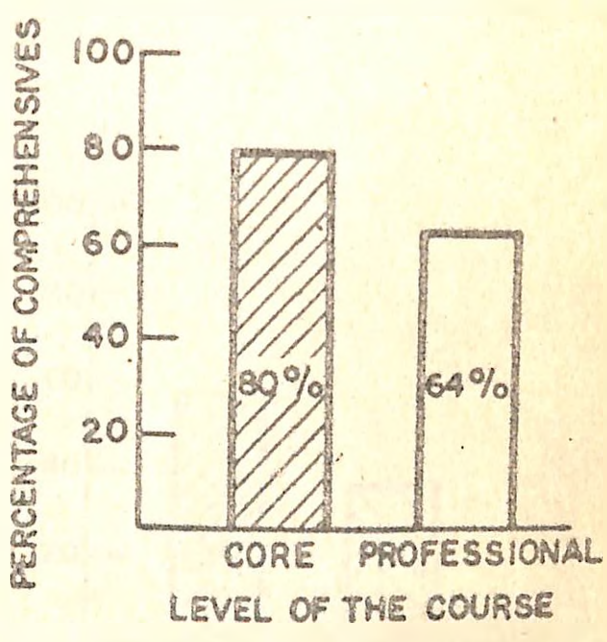


FIG. 3.60 PERCENTAGE OF COMPREHENSIVES WITH R ≥ 0.6.

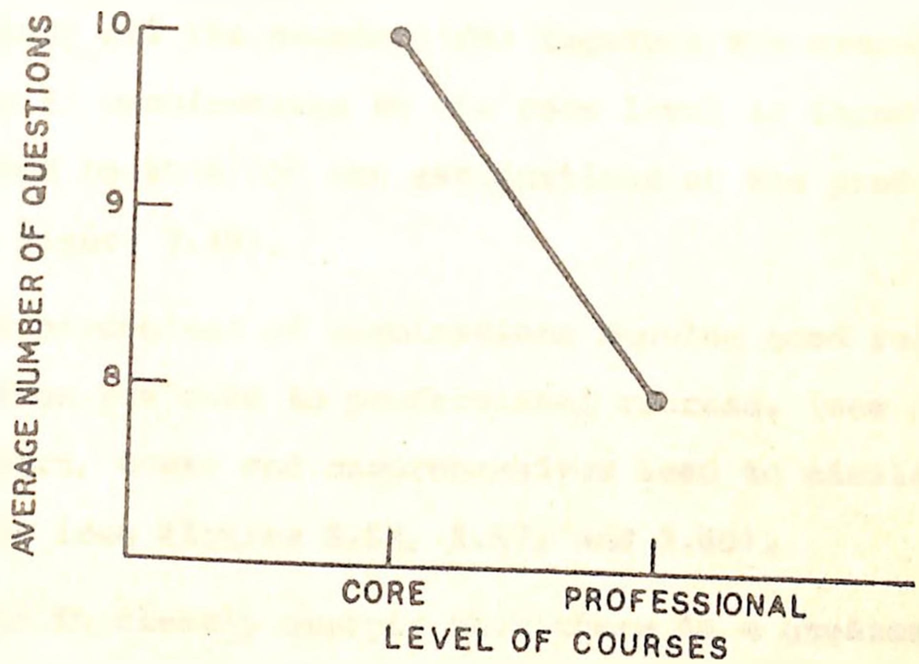


FIG. 3·61 AVERAGE NUMBER OF QUESTIONS IN EXAMINATIONS PLOTTED AGAINST LEVEL OF COURSES.

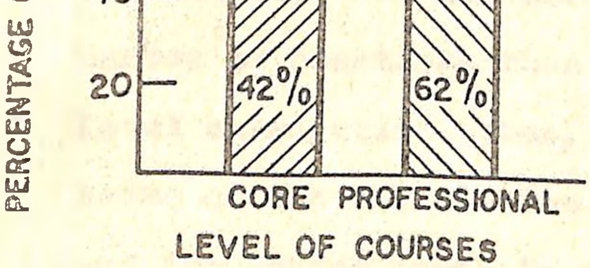


FIG. 3·62 PERCENTAGE OF EXAMINATIONS WITH NUMBER OF QUESTIONS ≤7.

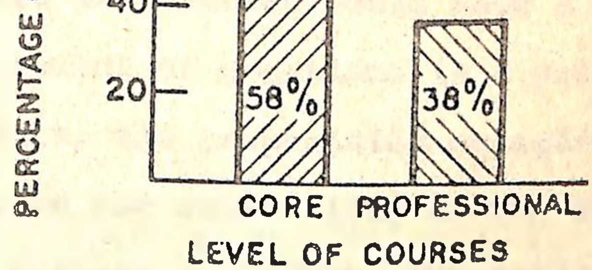


FIG. 3·63 PERCENTAGE OF EXAMINATIONS WITH NUMBER OF QUESTIONS >7.

examinations having questions equal to or less than 7 and more than 7, respectively.

(f) Summary of Observations

(1) Taking all the examinations together the average reliability of examinations at the core level is found to be more compared to that of the examinations at the professional level (see figure 3.49).

(2) The percentage of examinations showing good reliability decreases from the core to professional courses. (see figure 3.51). Quizzes, tests and comprehensives lead to similar observations (see figures 3.54, 3.57, and 3.60).

(3) Thus it clearly emerges that there is a greater probability that the examination at core level would have a better reliability than the examination at the professional level.

(4) The explaination for the above stated correlation between the examination reliability and the course level (in terms of the core or professional levels) may, as in the earlier cases, again be found in terms of the observation emerging from figures 3.61 so 3.63 that there is a greater probability that the core level examination would have a larger number of questions than the number of questions in a professional level examination. Thus, in turn, the proposition emerging in terms of the correlation between the reliability of the examination and its course level is also further confirming the earlier observed correlation between the reliability and the number of questions in an examination.

(vi) Reliability Observations According to Course Discipline

(a) Analysis for All the Examinations taken together

All the examinations selected for the study ~~have been~~ can be grouped as follows on the basis of their disciplines:

Table 3.25

Distribution of Examinations According to Disciplines of Courses

Total number of different type of Examinations \ Discipline	Science	Engineering	Humanities	Managemen
7 Quizzes	2	2	3	-
52 Tests	18	14	12	8
24 Comprehensives	8	4	7	5

The average reliability of examinations from various disciplines is given in figure 3.64. Examinations for Courses from Science, the humanities and management indicate almost the same average reliability (0.56, 0.55 and 0.56, respectively), where as it (the average reliability) is low in the case of Examinations from engineering courses (0.50). The percentages of examinations from various disciplines having reliability less than 0.5 and equal to or more than 0.5 are given in figures 3.65 and 3.66, respectively.

(b) Analysis for Quizzes

Quizzes studied were from science, engineering and the humanities. The distribution of quizzes in various disciplines is given in figure 3.67. Figures 3.68 and 3.69, for different disciplines, give percentages of quizzes having reliability less than 0.4 and equal to or more than 0.4, respectively.

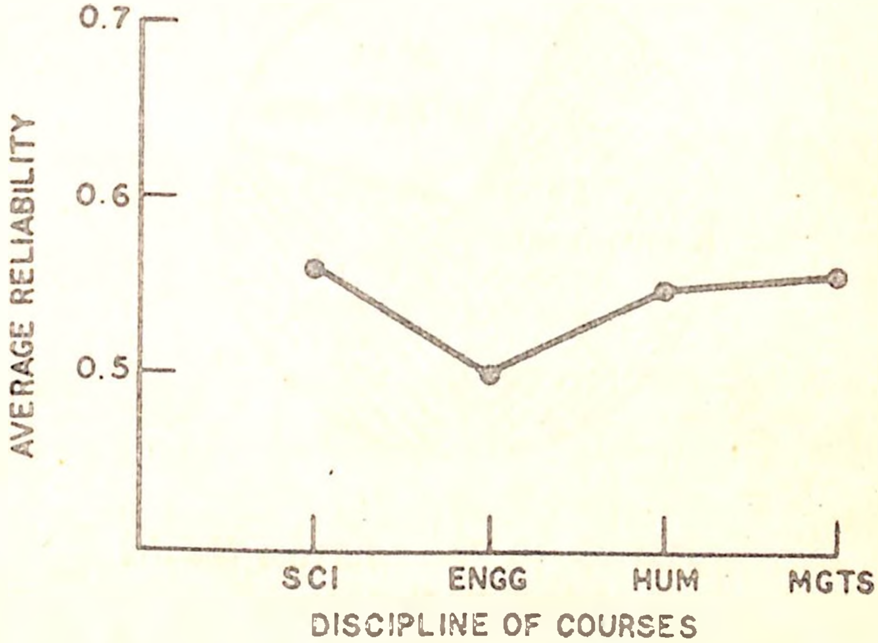


FIG.3·64 AVERAGE RELIABILITY OF EXAMINATIONS PLOTTED AGAINST DISCIPLINE OF COURSES.

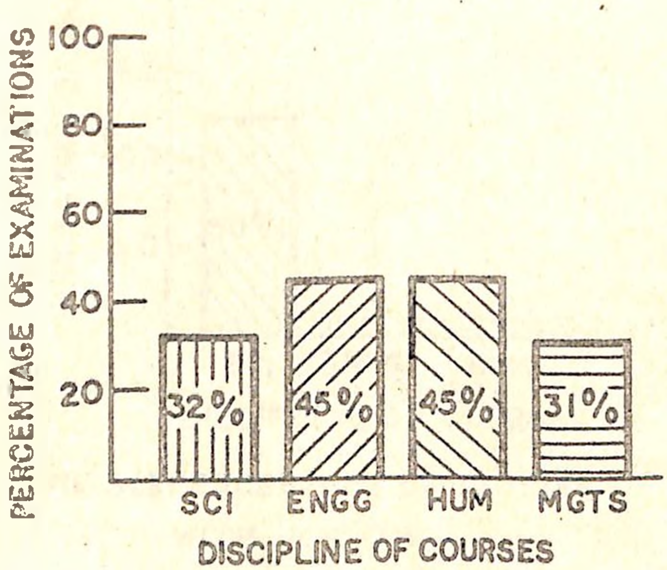


FIG.3·65 PERCENTAGE OF EXAMINATIONS WITH R<0.5.

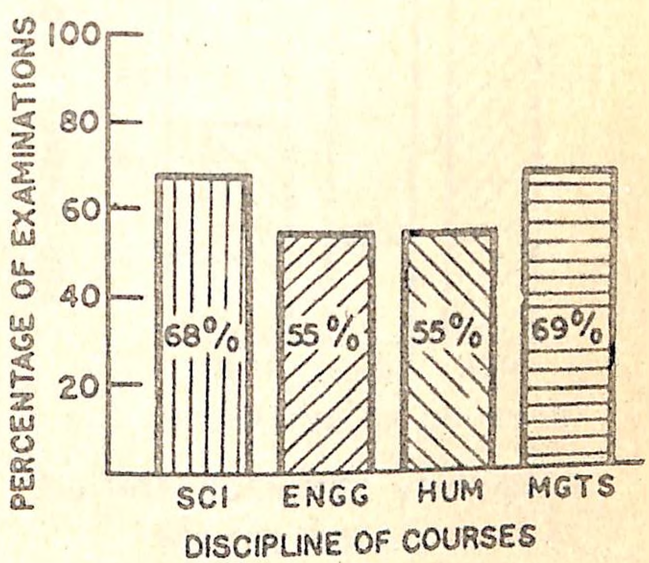


FIG. 3·66 PERCENTAGE OF EXAMINATIONS WITH R ≥ 0.5.

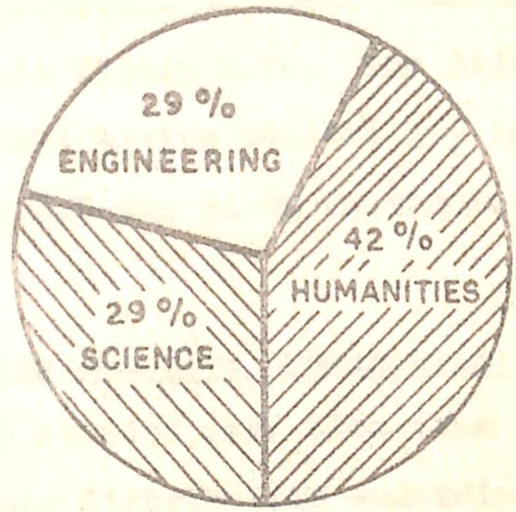


FIG. 3·67 DISTRIBUTION OF QUIZZES ACCORDING TO DISCIPLINE OF COURSES.

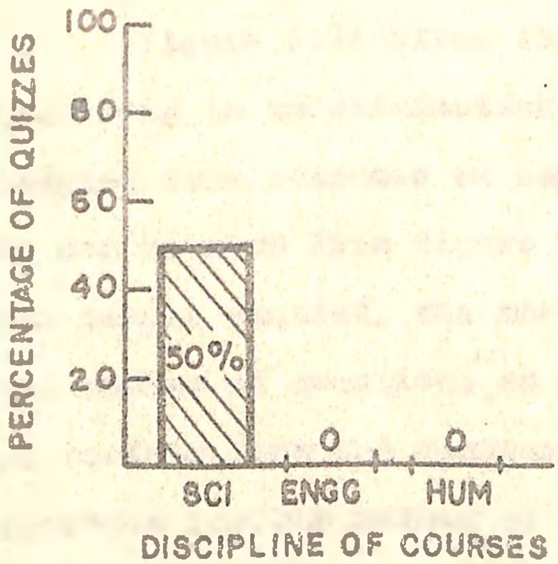


FIG. 3·68 PERCENTAGE OF QUIZZES WITH R < 0.4.

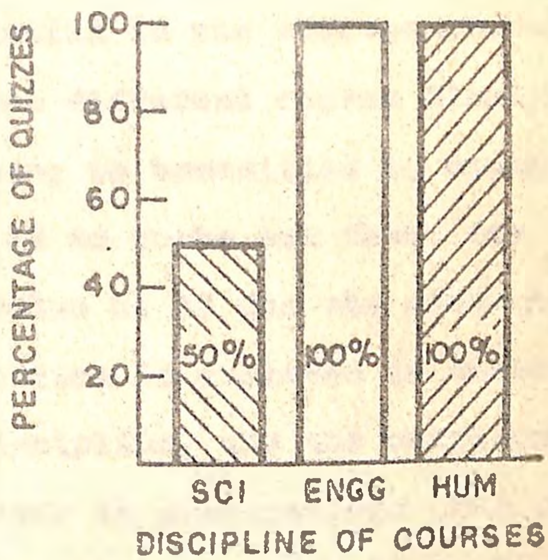


FIG. 3·69 PERCENTAGE OF QUIZZES WITH R ≥ 0.4.

(c) Analysis for Tests

The tests selected for the study are from all the disciplines. The distribution of tests according to their disciplines is shown in figure 3.70. For different disciplines, the percentages of tests having reliability less than 0.5 and equal to or more than 0.5 are given in figures 3.71 and 3.72 , respectively.

(d) Analysis for Comprehensive Examinations

Comprehensives studied were also from all the disciplines. Figure 3.73 shows their distribution according to disciplines of courses. For different disciplines, the percentages of comprehensives having reliability less than 0.6 and equal to or more than 0.6 are given in figures 3.74 and 3.75, respectively.

(e) Number of questions in an examination as a function of the Course Discipline

Figure 3.76 gives the variation in the average number of questions in an examination against different course disciplines ranging from sciences to engineering to humanities to management. As can be seen from figure 3.76, it so works out that, for the sample studied, the maximum value of 12 for the average of the number of questions in an examination is observed in terms of courses from the management discipline, and the corresponding averages for the number of questions in examinations from the science, engineering and humanities disciplines are 10, 6 and 8, respectively. Further, figures 3.77 and 3.78, for courses from different disciplines of sciences, engineering, humanities and management, give percentages of examinations having the number of questions per examination less than or equal to 7 and more than 7, respectively.

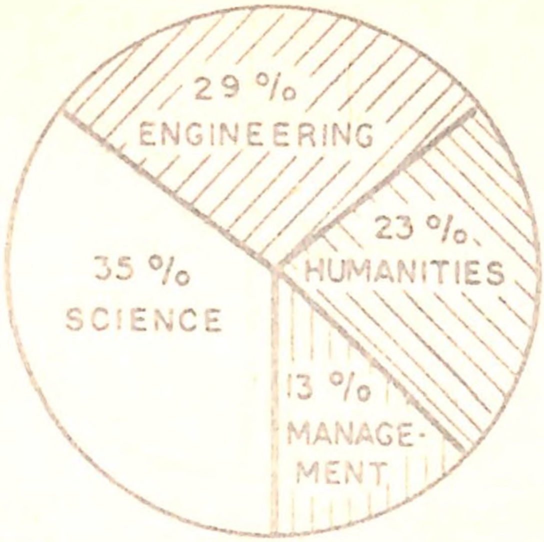


FIG. 3·70 DISTRIBUTION OF TESTS ACCORDING TO DISCIPLINE OF COURSES.

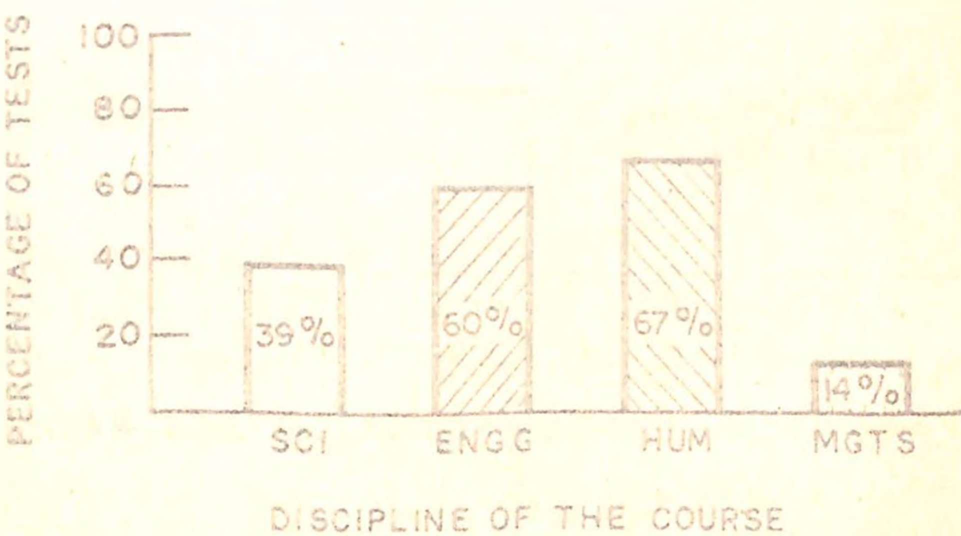


FIG 3·71 PERCENTAGE OF TESTS WITH R < 0.5.

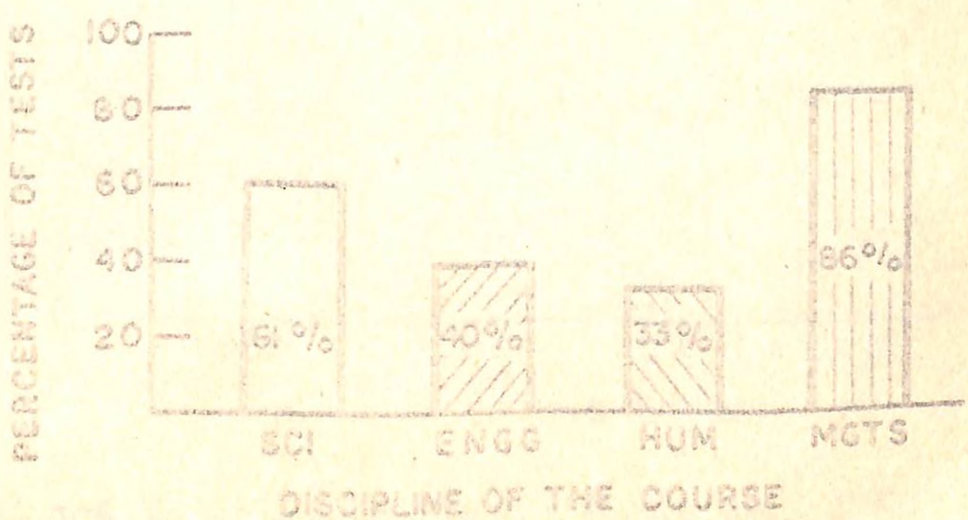


FIG 3·72 PERCENTAGE OF TESTS WITH $R \geq 0.5$.

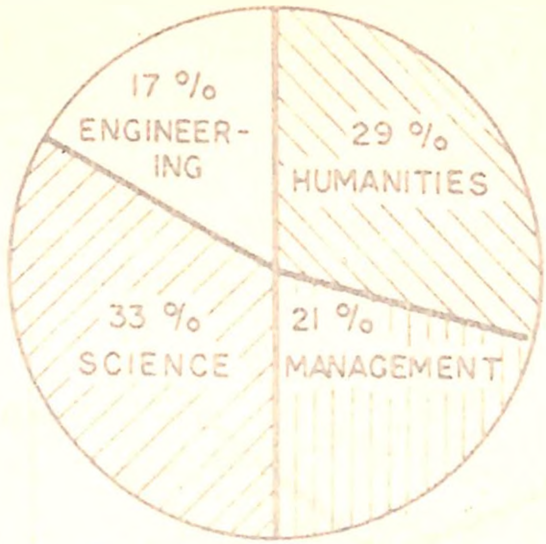


FIG. 3.73 DISTRIBUTION OF COMPREHENSIVES ACCORDING TO DISCIPLINE OF COURSES

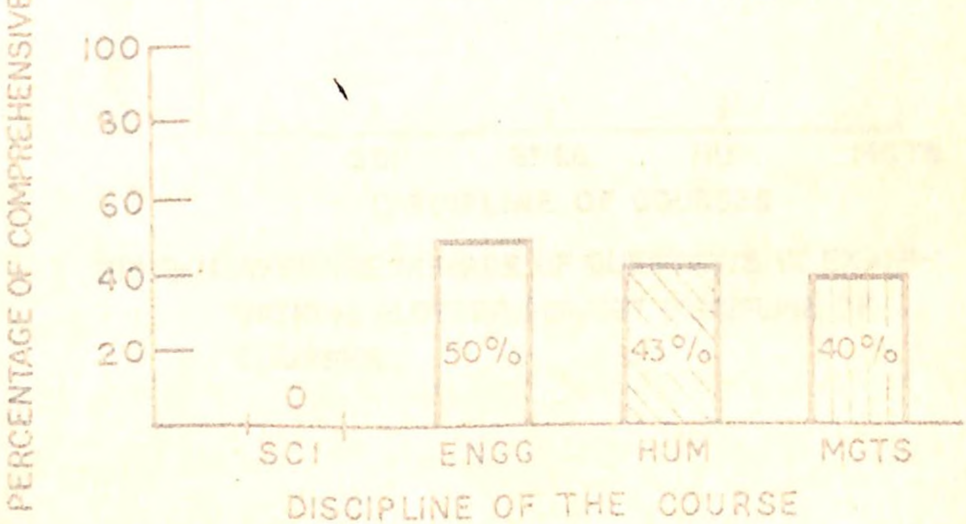


FIG 3.74 PERCENTAGE OF COMPREHENSIVES WITH R < 0.6.

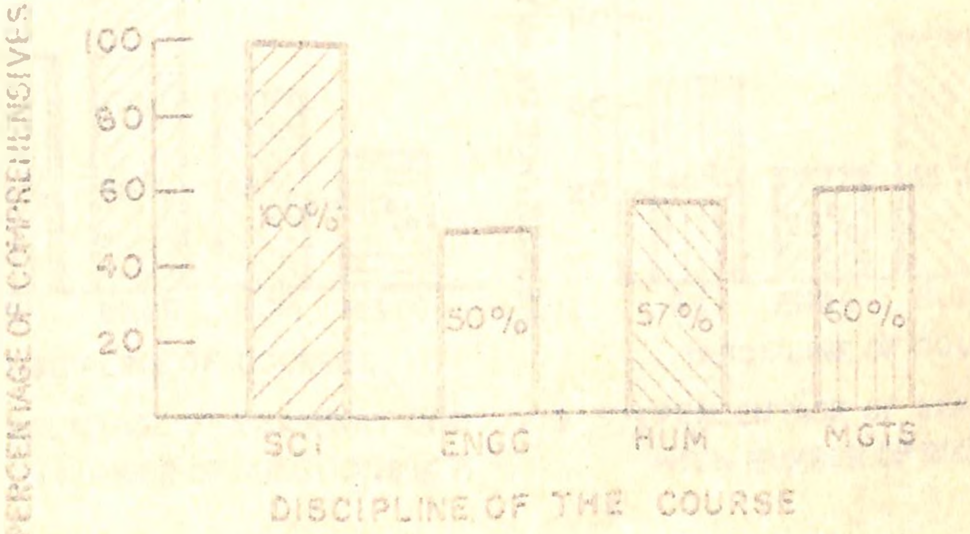


FIG 3.75 PERCENTAGE OF COMPREHENSIVES WITH R ≥ 0.6.

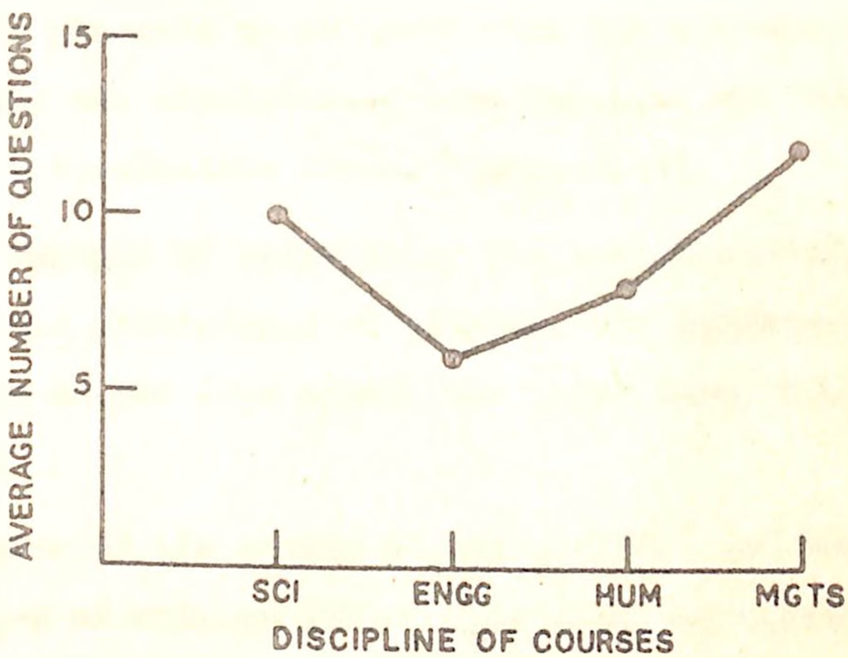


FIG. 3·76 AVERAGE NUMBER OF QUESTIONS IN EXAMINATIONS PLOTTED AGAINST DISCIPLINE OF COURSES.

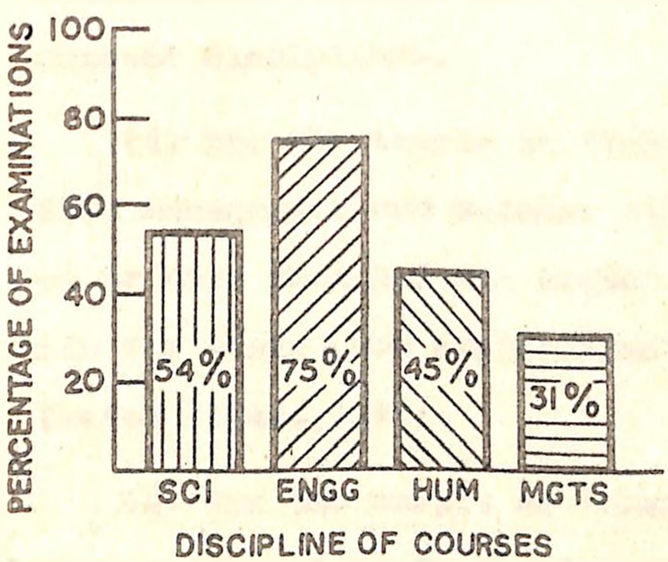


FIG. 3·77 PERCENTAGE OF EXAMINATIONS WITH NUMBER OF QUESTIONS ≤ 7.

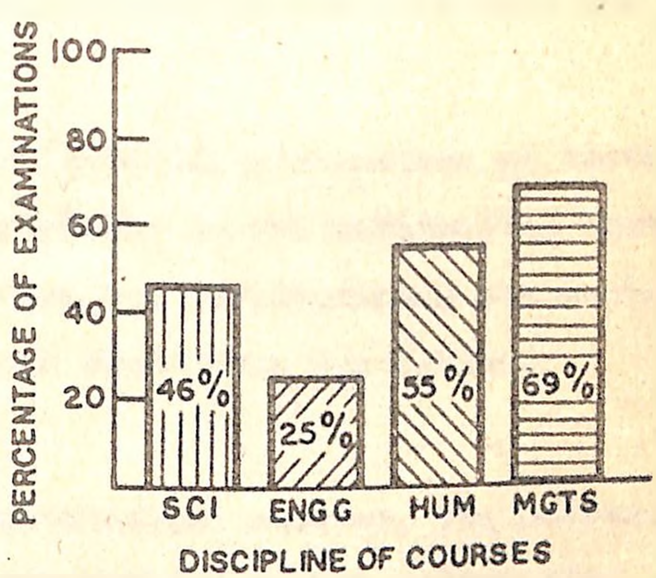


FIG. 3·78 PERCENTAGE OF EXAMINATIONS WITH NUMBER OF QUESTIONS > 7.

(f) Summary of Observations

(1) Within the sample of 83 examinations studied, the average reliability for examinations from engineering disciplines is observed to be lower than the averages of reliability for the examinations from science, the humanities and management disciplines (refer figure 3.64).

(2) The averages of reliability for the examinations for courses from the disciplines of science, the humanities and management are of the same order, the value being 0.56 (refer figure 3.64).

(3) In terms of the sample of quizzes that has been studied, the percentages of quizzes from engineering and humanities disciplines with reliability equal to or more than 0.4 are greater than the corresponding percentage for quizzes from the science disciplines (refer figure 3.69); thus, implying a greater possibility of a quiz from engineering or humanities disciplines having a better reliability than the quiz from the science disciplines.

(4) For the sample of 'tests' studied, percentages of tests from management and science disciplines having reliability equal to or more than 0.5 are higher than the corresponding percentages for the tests from engineering and humanities disciplines (refer figure 3.72).

(5) For the sample of 'comprehensives' studied, the percentage of comprehensives from science disciplines having reliability equal to or more than 0.6(the same being 100%) is higher than the corresponding percentage for management discipline (60%), is higher than the corresponding percentage for humanities

disciplines (57%), is higher than the corresponding percentage for engineering disciplines (50%).

(6) Finally, comparing figures 3.64 and 3.76, one can see that to a great extent the variations in the average values of reliability as also the average values of number of questions in an examination for examinations from various disciplines follow a similar pattern; thus, once again, confirming a good correlation between a reliability of an examination and the number of questions in it.

(vii) <u>Reliability Observations According to the Type of Questions</u>

For the purpose of the analysis in terms of the types of questions, all the examinations have been classified into following categories:

(i) Examinations with short answer questions of the descriptive type,

(ii) Examinations with short answer questions of the problem solving type,

(iii) Examinations with short answer questions of mixed type (i.e. containing descriptive as also problem solving questions),

(iv) Examination with long answer questions of descriptive type,

(v) Examinations with long answer questions of problem solving type, and

(vi) Examinations with long answer questions of mixed type.

As the names suggest, short answer and long answer questions would depend upon the length of the answer required. Descriptive questions are memory-oriented and require descriptive answers, where as problem-solving questions are application-oriented. In the mixed sub-category, examinations have a combination of different types of questions.

(a) Analysis for all the Examinations taken together

The average reliability of the examinations according to different categories of questions is given in figure 3.79. Examinations with short answer descriptive and short-answer problem solving type questions show an average reliability of 0.63 and 0.64, respectively. Examinations having long-answer descriptive, ~~as well as~~ long-answer problem-solving and mixed type questions indicate comparatively lower values of the averages of reliability, the same being of the order of 0.46, 0.46 and 0.49, respectively. For the various types of questions, the percentages of examinations having reliability less than 0.5 and equal to or more than 0.5 are given in figures 3.80 and 3.81, respectively.

(b) Analysis for Quizzes

As can be seen from figures 3.82 and 3.83, quizzes normally seem to have only short answer type questions. These questions could be of descriptive type or problem-solving type or mixed type.

(c) Analysis for tests:

As can be seen from figures 3.84 and 3.85, tests seem to have all the categories of the types of questions listed earlier. In specific terms for the various categories of questions, these figures give percentages of tests having reliability less than 0.5 and equal to or more than 0.5, respectively.

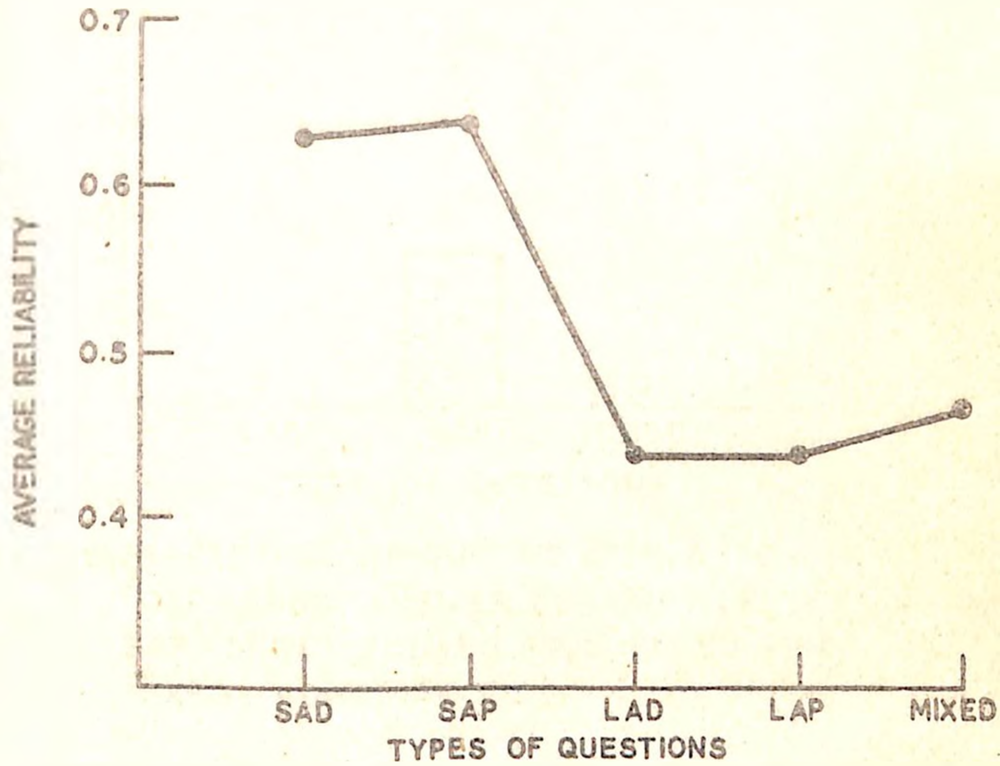


FIG. 3·79 AVERAGE RELIABILITY OF EXAMINATIONS PLOTTED AGAINST TYPES OF QUESTIONS (SAD - SHORT ANSWER DESCRIPTIVE; SAP - SHORT ANSWER PROBLEM SOLVING; LAD - LONG ANSWER DESCRIPTIVE; LAP - LONG ANSWER PROBLEM SOLVING; MIXED - COMBINATIONS).

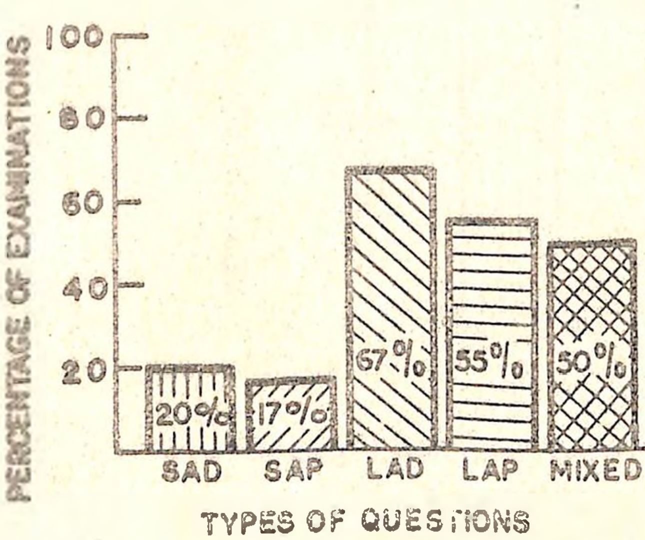


FIG. 3·80 PERCENTAGE OF EXAMINATIONS WITH R < 0.5.

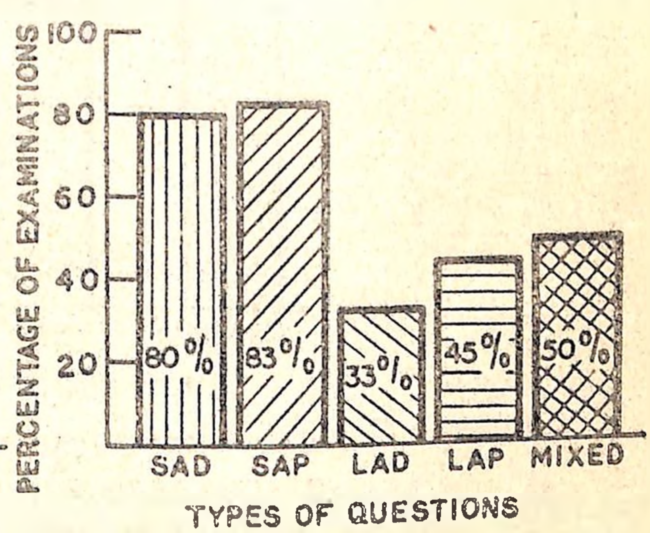


FIG. 3·81 PERCENTAGE OF EXAMINATIONS. WITH R ≥ 0.5.

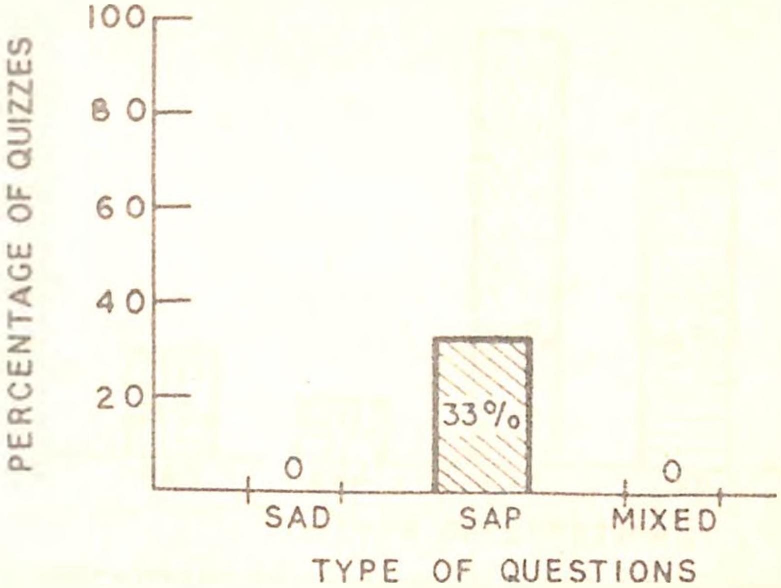


FIG. 3.82 PERCENTAGE OF QUIZZES WITH R < 0.4
(SAD - SHORT ANSWER DESCRIPTIVE, SAP - SHORT ANSWER PROBLEM SOLVING, MIXED - COMBINATIONS).

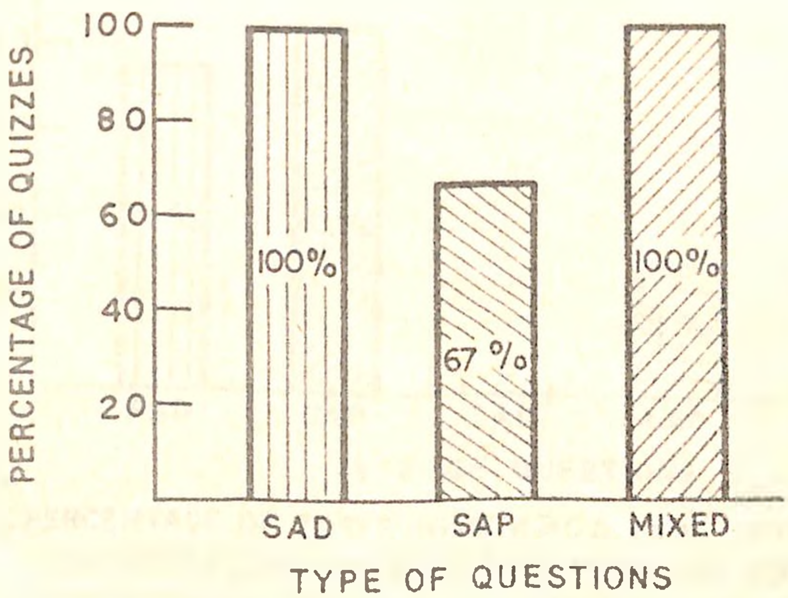


FIG. 3.83 PERCENTAGE OF QUIZZES WITH R ≥ 0.4
(SAD - SHORT ANSWER DESCRIPTIVE; SAP - SHORT ANSWER PROBLEM SOLVING; MIXED - COMBINATIONS).

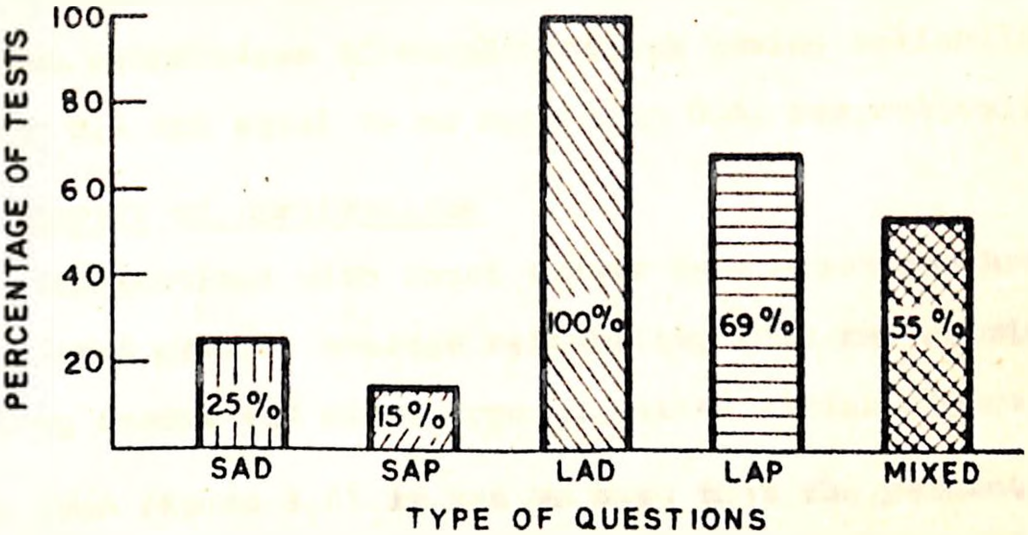


FIG 3.84 PERCENTAGE OF TESTS WITH R<0.5 (SAD- SHORT ANSWER DESCRIPTIVE; SAP-SHORT ANSWER PROBLEM SOLVING, LAD-LONG ANSWER DESCRIPTIVE; LAP-LONG ANSWER PROBLEM SOLVING; MIXED - COMBINATIONS).

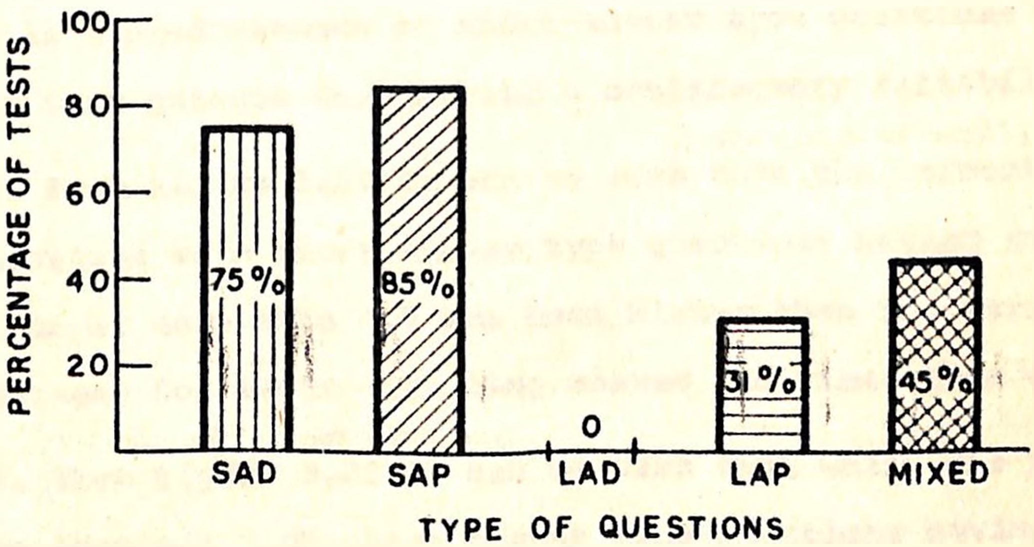


FIG 3.85 PERCENTAGE OF TESTS WITH R≥0.5 (SAD - SHORT ANSWER DESCRIPTIVE, SAP - SHORT ANSWER PROBLEM SOLVING, LAD-LONG ANSWER DESCRIPTIVE; LAP-LONG ANSWER PROBLEM SOLVING; MIXED - COMBINATIONS).

(d) Analysis for Comprehensives

For different categories of questions, figures 3.86 and 3.87 give percentages of comprehensives having reliability less than 0.6 and equal to or more than 0.6, respectively.

(e) Summary of Observations

1. Examinations with short answer type questions have demonstrated greater average reliability than the examinations with long answer and mixed type questions (refer figure 3.79).

2. From figure 3.81 it can be seen that the percentages of examinations with short answer type questions having reliability equal to or more than 0.5 are much higher than the corresponding percentages for the examinations with long answer type questions.

3. From figures 3.82 and 3.83 it can be observed that normally there is a predominance of short-answer type questions in quizzes and in turn quizzes demonstrate a satisfactory reliability level.

4. From figure 3.85 it can be seen that in case of tests, the percentages of examinations with short answer type questions having reliability equal to or more than 0.5 are much higher than the corresponding percentages for tests with long answer and mixed type questioons.

5. In case of comprehensives, from figure 3.86 it can be seen that while the percentages of examinations with short answer type questions having reliability equal to or more than 0.6 are invariably high, the corresponding percentage for examinations with long answer problem solving type questions is also high but so is not the case in terms of examinations having long answer descriptive type and mixed type questions.

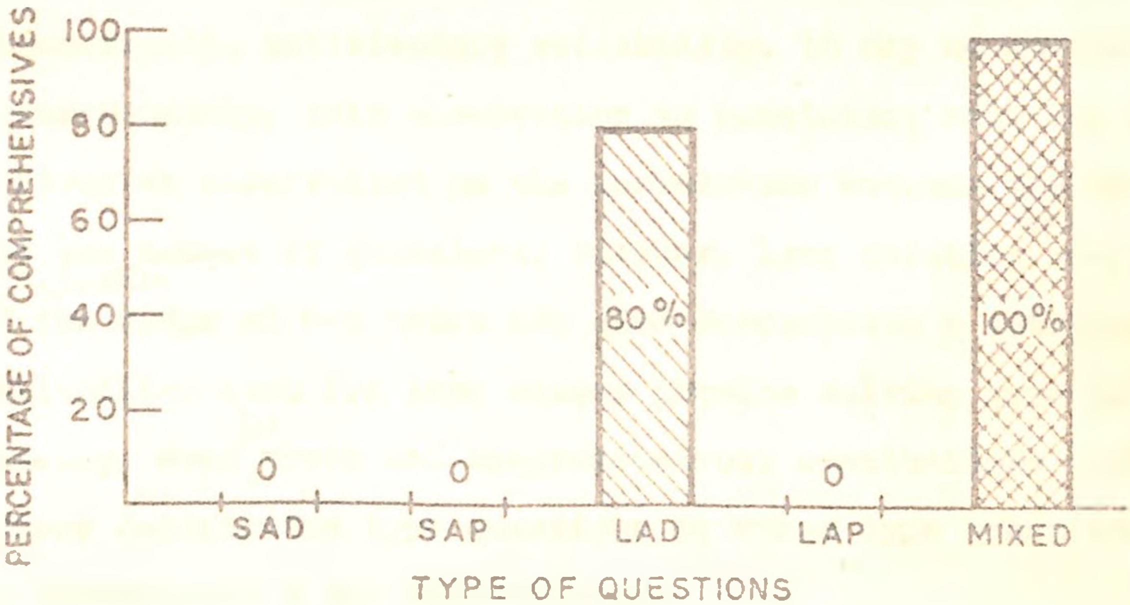


FIG.3·86 PERCENTAGE OF COMPREHENSIVES WITH R<0.6 (SAD-SHORT ANSWER DESCRIPTIVE; SAP-SHORT ANSWER PROBLEM SOLVING, LAD-LONG ANSWER DESCRIPTIVE; LAP-LONG ANSWER PROBLEM SOLVING; MIXED-COMBINATIONS).

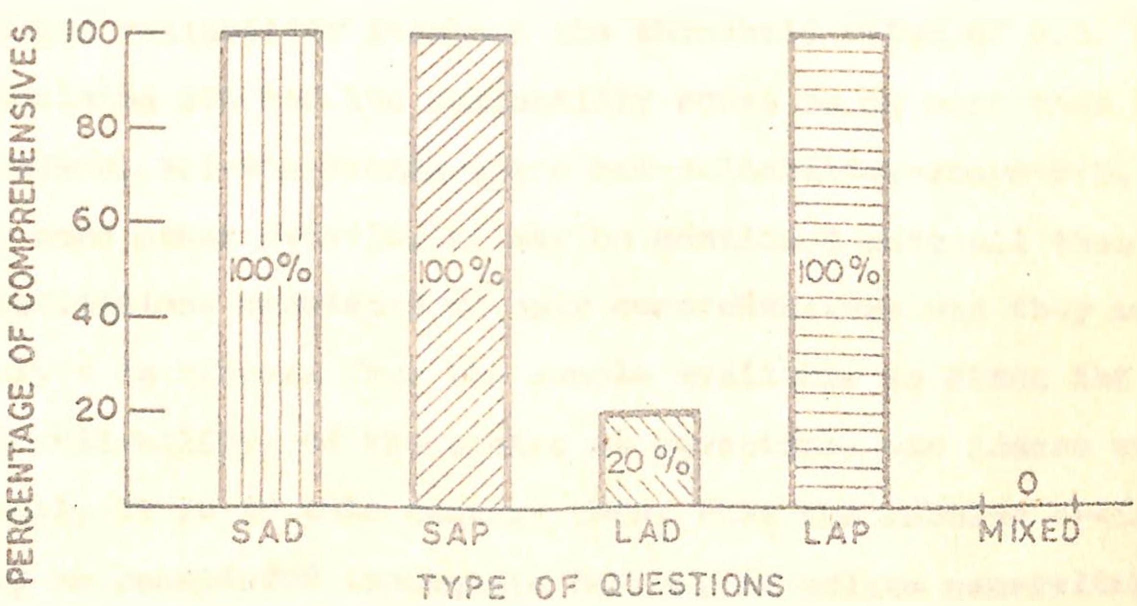


FIG. 3·87 PERCENTAGE OF COMPREHENSIVES WITH R ≥ 0.6 (SAD-SHORT ANSWER DESCRIPTIVE; SAP-SHORT ANSWER PROBLEM SOLVING, LAD-LONG ANSWER DESCRIPTIVE; LAP-LONG ANSWER PROBLEM SOLVING; MIXED-COMBINATIONS).

6. Thus, in summary, it emerges that examinations with short answer type questions have a greater probability of demonstrating satisfactory reliability. It may be mentioned that understandably, this observation is consistent with the earlier arrived at observation on the correlation between the reliability and the number of questions. Further, long duration examinations like comprehensives with duration of the order of 2-3 hours may also demonstrate a satisfactory reliability even for long answer problem solving type questions. Finally, even for tests and comprehensives, examinations with long answer descriptive type questions or mixed type questions seem to demonstrate a low reliability.

(viii) Reliability Observations According to the Choice of Questions in the Examinations

When the sample of examinations with choice of questions was analysed, as shown in figure 3.86, it emerged that 50% of it had reliability less than the threshold value of 0.6, while remaining 50% had the reliability equal to or more than 0.6. ~~Further, all the examinations had reliability above 0.5.~~ Coming to some other details, it may be mentioned that all these examinations consisted of only comprehensives and they were only 4 in number. Thus the sample available to study the effect on reliability, of the choice of questions, was indeed very small. It is in this context then, that the results available may be considered inadequate for any immediate generalisation in terms of the subject matter under discussion.

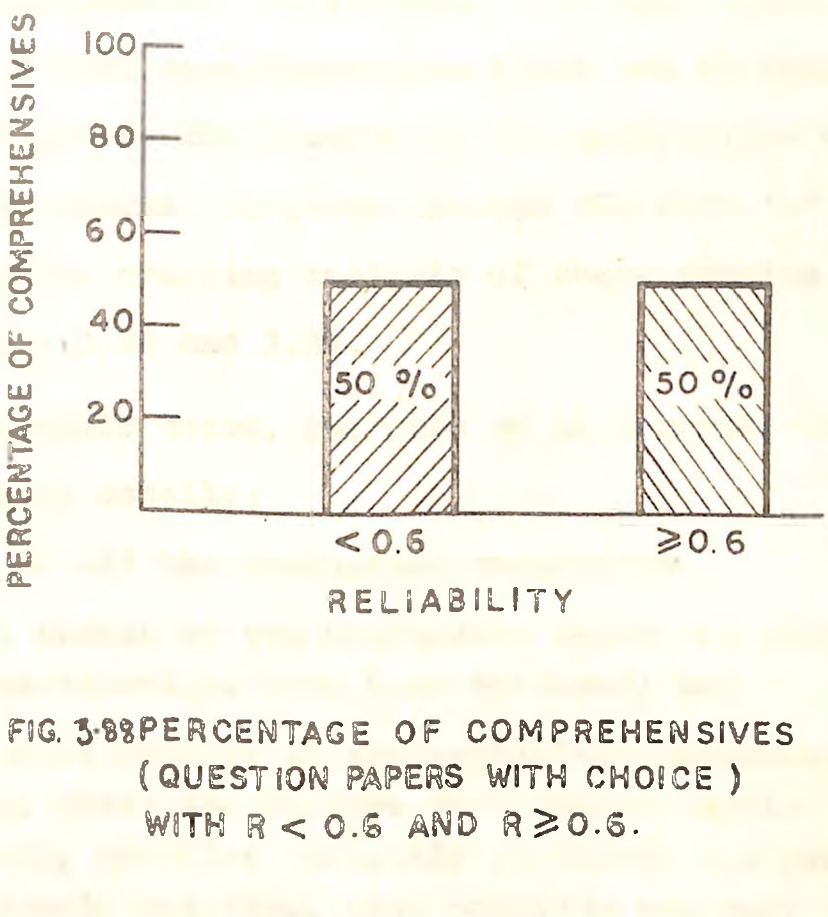


FIG. 3·88 PERCENTAGE OF COMPREHENSIVES
(QUESTION PAPERS WITH CHOICE)
WITH R < 0.6 AND R ≥ 0.6.

3.9.2 Analysis of the Results on the Correlations Between Various examinations of a Course:

This sub-section is devoted to the analysis of how the various components of evaluation for a given course are correlated. In another words, for different components of evaluation, this sub-section studies as to how well would a student, who has done satisfactorily in one component, performed in other components.

The correlation coefficients have been studied for in all 18 courses, 7 from core/foundation years and 11 from the professional years. The results on the correlation coefficients for these 18 courses are given through the Part 'b' of tables 3.4 to 3.22. The emerging analysis of these results is presented through tables 3.26 and 3.27.

In specific terms, for each of 18 courses, table 3.26 has given following details:

1. List of all the evaluation components,
2. Total number of the evaluation component pairs for which the correlations have been obtained, and
3. The exact numbers of the evaluation component pairs as coming under the various correlation levels, namely, slightly negative, slightly positive, low positive, moderately positive, high positive and very high positive. Further, each of these numbers is also presented as a percentage of the total evaluation component pairs studied for the course under consideration.

Against this, table 3.27 gives the analysis in terms of the 22 evaluation components pairs that could be generated through the data-base that was available with respect to the above indicated 18 courses. As can be seen from table 3.27, these

TABLE-3.26

Course-wise Analysis of the Numbers of the Evaluation-Component-Pairs as Coming Under Various Correlation Levels

Course Title	Listing of various Evaluation Components Incorporated	Numbers of the Evaluation-Component-Pairs as Coming under various correlation levels (Nos. in the bracket show the evaluation-component-pair No. as %age of the Total Evaluation-Component-Pair No. for the corresponding course)						Total No. of evaluation-component-pairs for the course under consideration
		-ve slight	+ve slight	+ve Low	+ve Moderate	+ve High	+ve Very High	
Core Courses								
Concepts in Science	Q1, Q2, T1, T2, T3, PR, C		6(28.57)	6 (28.57)	9 (42.86)			21
Physics I	Q, T1, T2, T3, C			2 (20)	8 (80)			10
Modern Physics	LQ, T1, T2, T3, LW, C			3 (20)	12 (80)			15
Probability &Statistics	Q1, Q2, Q3, Q4, Q5, T1		12 (33.33)	21(58.35)	2 (5.55)	1 (2.77)		36
Impact of Sc.&Tech.(1)	Q1, Q2, T1, T2, T3, C		2 (13.33)	9 (60)	4 (26.67)			15
Impact of Sc.&Tech.(2)	Q1, Q2, T1, T2, T3, C		2 (13.33)	5(33.34)	8(53.33)			15
Contemporary India				10(66.67)	5(33.33)			15
Total Evaluation-ComponentsPairs under Core Courses taken together			22(17.32)	56 (44.09)	48(37.80)	1 (0.79)		127
Professional Courses								
Solid State Physics	Q1, Q2, Q3, T1, T2, T3, HA, C		2(7.14)	10 (35.72)	14 (50)	2(7.14)		28
Bio-Physics	Q1, Q2, T1, T2, T3, C		1(6.67)	3(20)	10 (66.66)	1(6.67)		15
Combinational Math	T1, T, 2, T3, S, C			2(20)	6 (60)	2(20)		10
Control Systems Engg.	LQ, T1, T2, T3, LR, C			7 (46.67)	8 (53.33			15
Systems Modelling	Q, T1, T2, T3, C			5 (50)	5 (50)			10
Water & Waste Water Treatment	LQ, T1, T2, T3, C	1(10)	3(30)		1 (10)	4(40)	1 (10)	10
Intro. to Microprocessors	T1, T2, T3, C			1 (16.67)	5 (83.33)			06
Dev. Planning in India	Q1, Q2, T1, T2, T3, C	2 (13.33)	2 (13.33)	2 (13.33)	9 (60.01)			15
Systems Analysis (1)	Q, T1, T2, T3, PR, CA, CB			3 (14.28)	18 (85.72)			21
Systems Analysis (2)	Q, T1, T2, T3, C				10 (100)			10
Production Management	T1, T2, R1, R2, R3, R4, A1, A2, C	3 (8.33)	8 (22.22)	11 (30.55)	14 (38.9)			36
Total Evaluation-Component-Pairs Under Professional Courses taken together		6 (1.98)	16 (9.09)	44 (25)	100 (56.82)	9 (5.11)	1 (0.57)	176
Total Evaluation-Component-Pairs Under Core+Professional courses taken together		6 (1.98)	38 (12.54)	100 (33)	148 (48.85)	10 (3.3)	1 (0.33)	303

TABLE- 3.27

Evaluation Component - Category-wise Analysis of the Numbers of the Evaluation-Component-Pairs as Coming Under Various Correlation Levels

tegories of valuation-mponent-Pairs	Numbers of the Evaluation-Component-Pairs as coming Under Various Levels (Numbers in the Bracket show the Evaluation-Component-Pair Number As Percentage of the total Evaluation-Component-Pair Number for the Corresponding Evaluation-Component-Pair						Total Number of Evaluation-Component-Pairs for the Evaluation Component Pair under consideration
	-ve Slight	+ve Slight	+ve Low	+ve Moderate	+ve High	+ve Very High	
Q - Q		8(42.1%)	10(52.64%)	1(5.26%)			19(6.27%)
Q - P.R.		2(66.67%)	1(33.33%)				3(0.99%)
Q - H.A.			2(50%)	2(50%)			4(1.32%)
Q - T	2(2.94%)	6(8.82%)	35(51.48%)	25(36.76%)		1(11.11%)	69(22.44%)
Q - C		4(16.67%)	7(29.17%)	11(45.83%)	2(8.33%)		24(7.92%)
L.Q. - L.R.				2(100%)			2(0.66%)
L.Q. - T		1(11.11%)	3(33.33%)	4(44.45%)			8(2.97%)
L.Q. - C			1(33.34%)	1(33.33%)	1(33.33%)		3(0.99%)
P.R. - P.R.		2(33.34%)	2(33.33%)	2(33.33%)			6(1.98%)
P.R. - H.A.		3(37.5%)	3(37.5%)	2(25%)			8(2.64%)
P.R. - T	1(7.14%)	4(28.57%)	1(7.14%)	8(57.15%)			14(4.62%)
P.R. - C	1(14.29%)	2(28.57%)	1(14.29%)	3(42.85%)			7(2.31%)
L.R. - T			2(33.33%)	4(66.67%)			6(1.98%)
L.R. - C			1(50%)	1(50%)			2(0.66%)
H.A. - H.A.				1(100%)			1(0.33%)
H.A. - T			7(63.64%)	4(36.36%)			11(3.63%)
H.A. - C	1(33.34%)	1(33.33%)		1(33.33%)			3(0.99%)
T - T	1(1.92%)	1(1.92%)	13(25%)	34(65.38%)	3(5.78%)		52(17.16%)
T - C		4(7.14%)	10(17.86%)	38(67.86%)	4(7.14%)		56(18.49%)
T - S			1(33%)	2(67%)			3(0.99%)
C - C				1(100%)			1(0.33%)
C - S				1(100%)			1(0.33%)
All the Category of Pairs taken together	6(1.98%)	38(12.54%)	100(33%)	148(48.85%)	10(3.3%)	1(0.33%)	303(100%)

22 evaluation component pairs are contributed by different evaluation components such as class-room quizzes(Q), laboratory quizzes(L.Q.), laboratory work reports (L.R.), class-room project reports (P.R.), home assignments (H.A.), seminars (S), Tests (T), and, finally, comprehensives (C).

In specific terms, for each of the 22 evaluation-component-pairs, the table 3.27 gives following details:

(i) Total number of evaluation-component-pairs as coming under the particular evaluation component pair under consideration, and,

(ii) The exact number of evaluation-component-pairs from the total number of the evaluation-component-pairs for a given category as coming under each of the correlation levels, namely, slightly negative, slightly positive, low positive, moderate positive, high positive and very high positive. Further, each of these numbers is also presented as a percentage of the total number of the evaluation component pairs for the category under consideration.

Thus, it may be submitted that tables 3.26 and 3.27 within themselves present a self sufficient and therefore self-explanatory data-base with respect to the analysis details.

For the purpose of convenience, following abbreviations have been used in the present section.

TABLE 3.28

List of Abbreviations used in the present section

Full Name	Abbreviation
Quiz	Q
Tests	T
Comprehensive	C
Project report	P.R.
Home Assignment	H.A.
Lab. Quiz	L.Q.
Lab. Work	L.W.
Seminar	S

Summary of Observations:

Below are listed some of the major observations that are emerging from tables 3.26 and 3.27.

1. From table 3.27 it can be seen that from the sample studied as many as 98.02% evaluation-component-pairs have positive correlations, thus convincingly implying that the evaluation system on the whole is positively correlated having only 1.98% cases of negatively correlated evaluation-component-pairs.

2. Eighty one point eighty five percent of the sample of the evaluation-component-pairs studied have correlations within the levels positive low and positive moderate. Amongst all the levels, the positive moderate level has the maximum number of evaluation-component-pairs, this percentage being 48.85.

3. The sample of 303 evaluation-component-pairs studied has instances of pairs comering the entire spectrum of the coorelation levels from slightly negative to very high positive. The concerned percentages in the order of dominance are as follows:

Positive moderate - 48.85%; positive low 33%; slightly positive 12.5%; positive high 3.3%; negative slight 1.98%; and positive very high 0.33%.

4. Coming to the various evaluation component categories, it is noted that sample sizes for the categories namely, Q-T, T-C, T-T (i.e. within different tests), Q-C,Q-Q(i.e. within different quizzes), P.R.-T and H.A.-T were 68, 56, 52, 24, 19, 14 and 11, respectively. For all other evaluation component

categories the sample sizes were less than 9. Assuming the sample sizes equal to or more than 11 as effective ones, following observations emerge:

(i) The percentages of tests and comprehensives as also of different tests that are correlated in a moderately positive manner are as high as 67.86% and 65.38%, respectively. Thus, it can be said that tests and comprehensives as also one test to another ~~tests from~~ within a course ~~seem to be~~ are found to be well correlated.

(ii) The percentage of evaluation-component-pairs from the category of quizzes and tests as coming under the moderately positive correlation level is as low as 36.76%. Indeed, as much as 63.24% of the total sample size for Q-T comes under the categories from slightly negative to low positive; thus implying that quizzes and tests are not as well correlated as are T-C and T-T.

(iii) Consistent with the above observation, it also emerges that Q-C are also not as well correlated as are T-T and T-C.

(iv) One quiz to another within a course is found ~~Quizzes within themselves seem~~ to be very poorly correlated because as many as 94.74% of the evaluation-component-pairs corresponding to Q-Q are found to lie within the levels slight positive to low positive.

(v) Compared to quizzes, project report seems to show better correlations with tests in the sense that as many as 57.15% of evaluation component pairs from the category of P.R.-T are found to come under the moderately positive level. However, in terms of their correlation with comprehensive examinations, the P.R.-Q seem to compare better.

(vi) Finally, coming to the category of H.A.-T, it seems to emerge that they are not so well correlated, because as many as 63.64% of the total evaluation component pairs under this category are found to be under the positive low level.

5. Coming to table 3.26, it ~~seems to~~ emerges that, taking positive moderate level as a reference point, the evaluation component pairs from the professional year courses ~~seem to~~ show better correlations than the evaluation component pairs for the core level courses.

6. Further, instances of very high positive as well as slightly negative correlations are visible only in case of the professional level courses; thus implying examination components at the professional level are ~~seem to be~~ more prone to demonstrate extreme cases of correlation levels.

7. Within the entire spectrum of the 18 courses studied only one course i.e. Systems Analysis offered in I semester, 1978-79, demonstrated all the correlations within the positive moderate level, while all other courses necessarily had correlations corresponding to levels lower then positive moderate. The data base with respect to this course was collected on two occasions; the analysis for the one occasion being mentioned above. On the second occasion i.e. during its offering in I semester 1977-78. also for this course, as many as 85.72% of the evaluation component pairs come under the category of positive~~ly~~ moderate level; the remaining 14.28% having corresponded to the low positive level; thus signifying a consistency in terms of the results of the analysis. In a futuristic context it would be worthwhile to

explore the factors leading to such a satisfactory correlations between the various evaluation components of this course so that the improvements can then be brought elsewhere based on the lessons emerging.

8. Other courses for which evaluation components have demonstrated fairly good correlation are Physics I, Modern Physics and Introduction to Microprocessors. All these three courses have correlations only between positive low and positive moderate levels, the percentages of the evaluation components pairs under positive moderate level being equal to or more than 80%.

9. Thus, on the whole it can be said that, while, the evaluation component-pairs are positively correlated, the bulk of them seem to lie within positive moderate and positive low levels. Further, tests and comprehensives seem to show best correlation. It will be worthwhile to pursue how the level of correlation between tests and comprehensives can be further increased and how this effort can be enlarged to include the larger variety of evaluation components beyond the categories of tests and comprehensives.

3.9.3 Analysis of Results on Course Reliability

This sub-section gives an analysis of the course reliability of 17 courses selected for the study. Table ~~3.2.3~~ 3.23 gives the list of these courses and their course reliability coefficients denoted by Rc. It is very interesting to note that the course reliability coefficients are quite high. Almost all the coefficients are greater than 0.7, except in one case, where also the course reliability is as high as 0.66. Consistent with the earlier observations vis-a-vis the correlation between the examination reliability and the number of questions in an examination, in ~~this~~ case of course reliability also it is noted that an average number of evaluation components for the courses studied is 6, which is reasonably high. This in turn then provides an explanation for the high course reliability observed for the courses under consideration.

The practice of conducting higher number of evaluation components for assessing the student is the core of the continuous evaluation system. Thus, this system can then also be viewed to have the advantage of a good course reliability.

Apart from the parameter of a large number of evaluation components, there can also be another factors in the above context, contributing to the high course reliability. For example, as one is aware, in a continuous evaluation system, a course evaluation consists of several evaluation components and each of the evaluation components consists of several questions. The course reliability is determined on the basis of the marks obtained by students in these different evaluation components.

Thus, each evaluation component represents a composite examination which samples over a fairly wide area of the subject matter. This then should be reducing the randomness in the evaluation, thus consequently increasing the course-reliability.

3.10 Conclusion

This chapter has basically dealt with the problem of the test analysis for the examinations under the internal system of continuous evaluation at the tertiary level. Towards this, the chapter has in detail analysed 19 selected courses from the semester-wise offerings at BITS, Pilani. For the purpose of the totality, these courses have been drawn from the entire spectrum of disciplines and levels of the system of higher education. In specific terms, for the examinations as coming under the so selected courses, the chapter has dealt with different analytical aspects, namely, the reliability of examinations, inter-correlations between various examinations and the course reliability.

Thus, the chapter has studied coefficients of reliability for in all 83 examinations, consisting of 7 quizzes, 52 tests and 24 comprehensives. These reliability coefficients for the different evaluation components have been in detail analysed against various parameters such (i) the number of questions in an examination, (ii) number of students attending the examination, (iii) duration of the examination, (iv) level of the course, (v) discipline of the course, (vi) type of examination (i.e. long-answer, short-answer, etc.), (vii) category of the course in terms of the core course or the professional course and, (viii), finally, the aspect of the choice in an examination. Further for various evaluation components such as quizzes, tests, viva, seminar, project-reports, home-assignments, lab-quiz, lab-work, comprehensives, etc, as reflected through the above stated course the chapter has also studied the theme of intercorrelations between the various examinations of a course. Thus, the chapter has analysed in all 303 evaluation-component-pairs for their intercorrelations, 127 corresponding to the courses from the core level and 176 corresponding to the courses from the professional level. And, as regards to the theme of the course-reliability, consistent with the constraint of the availability of the proper data, the same has been studied for in all 17 courses from the total list of 19, selected for the overall investigation as undertaken in this chapter.

Below are briefly listed some of the main observations as they have emerged from the above study:

1. Reliability is affected by the number of questions given in an examination. In specific terms, larger the number of questions in an examination, higher is the reliability.
2. The number of students is also a factor which influences reliability. Broadly speaking, larger the number of students (in an examination), better is the reliability. At this stage, it may be mentioned that, normally, examinations for courses attended by a large number of students are observed to be characterized by large number of questions.
3. Longer the duration of an examination, greater the possibility of its reliability being higher. Interestingly the longer duration examinations are also, normally, characterized by the larger number of questions.
4. Normally, the average reliability of an examination decreases as one moves from say courses at the I and II

year levels to the courses at the IV and V year levels. In this case, also it may be noted that, on an average, the examinations for the courses at the lower level have more number of questions than that (i.e. number of questions) for the examinations for the courses at the higher level.

5. On the whole, examinations for the courses from the Science, Humanities and Management disciplines have demonstrated better reliability as compared to the examinations for the courses from the ~~humanities and~~ engineering disciplines. In this context, it may be mentioned that, normally, the examinations from the disciplines of science and management have been observed to contain more number of questions than the humanities and engineering examinations, in that order.

6. It emerges that the examinations with short answer type questions have a greater possibility of demonstrating satisfactory reliability.

7. On an average, the reliability of examinations in core courses is observed to be better than that for examinations from professional courses. In this case, too, it is noted that, normally, examinations in core courses have larger number of questions than that for the examinations from professional courses.

8. There are only four examinations, that too comprehensives, in which choice in questions was provided. Out of these four, for two the reliability is less than the threshold value of 0.6, while for the remaining two the reliability is equal to or greater than 0.6. As the sample size is too small (i.e. 4 examinations only), it is difficult to make any generalization as to the effect of choice on the reliability. However, one can infer from the data that teachers on the whole, do not encourage choice in an examination.

9. Out of the 303 evaluation-component-pairs studied for their intercorrelations, 98.02% evaluation-component-pairs demonstrated positive correlations. Further, 48.85% pairs

showed positive moderate correlation, while 33% demonstrated positive low correlations; thus implying that as many as 81.85% of the evaluation-component-pairs had correlations lying between the levels of positive low and positive moderate.

10. It is observed that tests and comprehensives as also one test to another within a course are well correlated.

11. Quiz to another quiz within a course showed a poor correlation. Further, quizzes also demonstrated poor correlations with tests and comprehensives for a course.

12. All the courses selected for the study demonstrated high course reliability.

Summing up, it can be said that the high reliability is obtained either in the case of a course, comprehensive, test or quiz, etc. provided the components of the evaluation instrument are sufficiently large and the situation in which evaluation is done ensures the consistency in performance due to mutually reinforcing components which is typical of any internal continuous evaluation system. This, of course, has been reflected in variety of ways while studying the reliability in terms of various parameters discussed above.

Finally, the above analysis of reliability provides a methodology, as well as direction in which some future studies can be undertaken so as to develop a data base for periodical monitoring of an evaluation system and giving feed-back to the teachers in a typical continuous internal evaluation system. In a futuristic manner, for the courses, where reliabilities and intercorrelations have demonstrated extreme values, it is suggested that further in depth studies can be taken up.

Above then is a comprehensive statement of the conclusions as emerging from the test analysis study as undertaken through this chapter. With this then the thesis switches over to another important aspect of the examination study, namely, 'Validity', pursued through the next chapter.

CHAPTER - 4

VALIDITY ANALYSIS WITH PARTICULAR REFERENCE TO THE CRITERION-RELATED VALIDITY

4.1 Introduction

In Chapter 3, examinations have been analysed in terms of their reliability. Another important parameter for assessing the quality of an examination is its validity. Literature defines various kinds of validity studies that can be conducted on an examination. This chapter studies the examinations in terms of 'criterion-related validity'.

In general, most of the examinations have more or less similar educational objectives. With this assumption, the question arises, how far the performance in one examination relates to or predicts the performance of some other concurrent or future examination. The correlation between such examinations of similar objectives is a measure of criterion-related validity. In this chapter an attempt has been made to study the criterion-related validity of the examinations by analysing the correlations between the performance of students in various examinations.

A brief description of the validity definitions and types of validity is given in section 4.2. Section 4.3 gives a review of studies on validity. The objective, assumptions and statement of the study are given in section 4.4. The technique for the proposed validity study has been discussed in section 4.5. Section 4.6 gives the sample and data collection procedure. Finally, the results, analysis of results and the conclusions have been presented in sections 4.7, 4.8 and 4.9, respectively.

4.2 Validity Definitions

4.2.1 Broad Definition

Perhaps the commonest definition given for the term validity is that it refers to the extent to which a test measures what it is supposed to measure (Anastasi (1976), Ebel (1979), Dick & Hagerty (1971), Gronlund (1976), Popham (1975)); i.e., an examination which is supposed to measure the decision making ability of students, would not be valid unless it really measures the decision-making ability.

Reliability refers to the consistency of measurement irrespective of whatever is being measured. In case of validity, the basic question is 'what is being measured'. Validity always refers to some particular objective. An examination may be valid for decisionmaking ability but may not be of much use for assessing say, memory of students. Thus, the validity of an examination is analysed with reference to the declared purpose of the examination.

4.2.2 Types of Validity and their Definitions

American Psychological Association (1974) has identified three basic types of validity

(i) Content validity

(ii) Criterion-related validity

(iii) Construct validity

(i) Content Validity

Various researchers have given definitions for the above categories of validity. Some of the definitions for content validity are as follows :

"Content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured". (Anastasi, 1976)

"The extent to which the items in a test do, infact, sample an area is termed the content validity of the test". (Dick & Hagerty, 1971)

".....(Classroom tests) are supposed to sample representatively and adequately the content of the course of instruction". (Ebel, 1979)

"The extent to which a test measures a representative sample of the subject-matter content and the behavioural changes under consideration". (Gronlund, 1976)

Thus, it can be said that the efforts to define content validity by various researchers are same. The procedure for establishing content validity can be summarised as follows (Gronlund 1976).

While planning the course work, one decides the subject matter, various topics to be covered in it and, finally, the objectives in terms of behavioural changes which are to be achieved through the contents of the course. Later, in the process of education, outcomes of the learning are evaluated. This is generally achieved by giving examinations to the students. The question of content validity of examination arises at this stage. To know whether the examination has content validity, the content and objectives of the examination paper are matched with the content and objectives of the course work. The extent to which these two match, the examination would be said to have content validity. The thesis has studied the content validity of practice school evaluation at BITS which would be discussed in detail in chapter 6.

(iii) Criterion-Related Validity

Gronlund (1976) defines criterion-related validity as "the extent to which test performance is related to some other valued measure of performance". Lindzey and Aronson (1968) explain, "whenever there is an identifiable criterion variable for which a test purports to measure, the test may be used to estimate an individual's present standing on that variable or to predict his future standing". And Popham (1975) writes, "when we employ a criterion-related validation approach we attempt to correlate performance on a measure (the one we are hoping to validate) with an independent external criterion".

There are, of course, several other researchers who have defined criterion-related validity but they all bring the same meaning to it.

Criterion-related validity is of importance whenever the examination marks are to be used to predict future performance or to estimate the present performance. Hence, it has been further classified as

a. Predictive validity

b. Concurrent validity

The predictive validity refers to the extent to which examination scores predict some future performance, whereas the concurrent validity relates the performance in one examination to the performance in a contemporary examination. Both predictive and concurrent validity are represented by coefficient of validity. The coefficient of validity is nothing but the correlation between scores obtained in two appropriately identified examinations. Validity coefficient is interpreted in the same way as coeffient of correlation. It may be remarked that in the interpretation of the validity the negative coefficient of correlations, if any, are ignored. (Gronlund, 1976). The present chapter studies examinations in terms of criterion-related validity.

4.3 Review of Literature on Validity Studies

Most of the studies on validity of examinations are confined to predictive validity. A brief summary of these studies is given below:

4.3.1 Studies on Validity Outside India

Wiseman (1956) in a sample of 117 students studied the criterion-related validity of entry test scores including an essay test using the following criteria

1. Total school certificate result
2. School certificate English language
3. School certificate English literature
4. Teachers overall estimate
5. Teachers estimate of written English

Correlation were calculated between different sets; and it was found that the addition of essay tests increased the coefficient, whereas the value of coefficient was lowered when essay test scores were excluded from entry test scores.

Pidgeon and Yates (1969) in a study of 473 students who had completed their secondary school course found that the essay type examination in English had less predictive validity for success in secondary schools than the results obtained by objective type examinations in English.

Hudson (1960) explored predictive validity of examinations for predicting eminence in scientific research. He concluded that the division or degree was not valid criterion for predicting eminence in scientific research.

Coffman et. al. (1966) studied the validity of essay type test for predicting writing ability. They selected XI and XII standard students from 24 American secondary schools. Each student was subjected to six objective type, two semi-objective type and five essay type examinations. Each student was supposed to write on five different topics. Scores obtained by students in essay test were considered as criterion score for predicting writing ability. Validity coefficient of essay type examination ranged from 0.33 to 0.35.

Coffman et. al. also observed that validity of examinations was lowered if choice was given. In such cases student's marking depended more on the type of topic selected by student than how well he had written.

Bracht and Hopkins (1968) found that CGPA correlated very well with performance in essay and objective tests but objective testscorrelated better than essay tests. This observation is supported by several other studies conducted abroad as well as in India.

Holloway et. al (1967) studied the validity of essay and viva voce examination. They found that the essay marking was influenced more by style and presentation than by factual content. They also observed that personality was an important factor to get success in viva voce.

4.3.2 Studies on Validity in India

There are several studies on the validity of examinations for predicting future performance. These studies have been conducted at college as well as university levels. Some of the well known studies are given below .

Taylor (1962) observed following correlations between various groups at the college level :

TABLE 4.1

Correlation coefficients between various examinations

Examinations	Coefficient of correlation
Physics theory with physics practical	0.14
Chemistry theory with chemistry practical	0.14
Physics theory with chemistry theory	0.30

Taylor concluded that the marks were not valid measures at all. His conclusions are based on the assumption that if marks are the valid measures of achievement then there should be good correlation between the performance in various subjects.

Patel (1962) reported moderate correlation (0.62) between the performance in practice teaching and B.Ed practical examination. He also studied the correlation between the performance in part I and part II examinations. The correlation was found to be 0.63. Like Taylor's study, Patel's study was also subjected to the assumption that the same ability is being tested by various examinations.

Kamat and Deshmukh (1961) observed following correlations between various examinations for a group of students

TABLE 4.2

Correlation coefficients between various examinations

	S.Sc	Inter Arts	Inter Science	B.A.	B.Sc.
S.S.C		0.60	0.61	0.49	0.21
Inter Arts				0.63	
Inter Science					0.42

Misra, G.S. (1964) reported following correlations :

TABLE 4.3

Correlation coefficients between various examinations

	H.S.	Intermediate	I.A.	I.Sc.	I.Sc. Ag.
Intermediate	0.71				
I.A.	0.73				
I.Sc	0.84				
I.Sc. Ag.	0.72				
Graduate	0.83	0.87			
B.A.	0.69		0.69		
B.Sc.	0.85			0.87	
B.Sc. Ag.	0.75				0.76

As can be seen from the above table Correlations between ~~various examinations~~ performances in various examinations came out quite good; hence he concluded that the results of high school and intermediate examinations are good predictor for predicting performance at university level.

Lele et. al (1963) reported 0.77 coefficient between S.S.C (English) and P.Sc (English) performance of M.S. University Baroda.

Taylor, Tluanga and Misra (1966) correlated the matriculation and P.U. examinations and the value of correlation was found to be 0.72.

Harper (1963) correlated the marks obtained in an objective examination of knowledge in English poetry with an essay examination in the English poetry in several intermediate colleges in U.P. The validity coefficient ranged from - 0.25 to 0.82. He concluded that the difference in validity coefficients may be either due to inherent differences between the institutions or may be due to low reliability of essay examinations.

Harper (1963b) in some other study of same nature studied whether objective test or essay test would be better predictor for performance in an essay examination. He selected 72 intermediate students and arrived at a conclusion that the objective type tests were better predictors than essay tests for predicting performances in the same length of essay tests.

Misra (1976) has discussed that using examination-marks for predicting performance in real life is more important than predicting performance in some other examinations. With this objective he studied the validity of teacher's training programme for predicting performance in real life. He selected 118 teachers of Boy's Intermediate College of Allahabad and investigated how far their teaching had improved by undergoing a teacher's training course. He concluded that the training of the teachers had not improved teacher's teaching efficiency significantly.

These are then some of the major studies on validity of examinations conducted in India. A detailed review of these studies in India as well as abroad is given by Harper & Misra (1976). As one goes through the above details, what emerges clearly is that, normally, the studies have only been on predictive validity. It is also found that most of the studies are at the college level. Studies at university level are few in number and almost nil for engineering education. Also one does not find any study on validity of examinations for the internal system of continuous evaluation characterised by the semester pattern.

4.4 Objective, Assumptions and the Statement of the Present Study

It is against the above frame work that this thesis proposes to apply itself to the study of the validity of examinations for the internal system of continuous evaluation characterised by the semester pattern. More specifically, if chapter 6 discusses, the content ~~and criterion related~~ validity vis-a-vis the practice school system of education, this chapter is devoted to study the subject matter of the criterion-related validity for the campus based examinations in the above context. The system at BITS provides a very convenient and realistic case study for this research. As indicated in chapter (1), this system has following main features:

1. Integrated broad-based education,
2. Relevance to social and national needs,
3. Strongly emphasises the overall development of the student's personality,

4. Semester patterns of course offering,

5. Evaluation is totally internal and relative,

6. Continuous with feedback to the students,

7. Final performance measured in letter grades in each course, each semester,

8. Total performance measured in terms of CGPA,

9. No fail grade, and

10. All answer papers, inclusive of comprehensive, shown to studen

As has been mentioned in section 4.2, the criterion-related validity is of importance whenever the examination marks are to be used to predict future performance or to estimate present performance. The technique for estimating the criterion-related validity would be discussed in the next section. However, it should be pointed out that the two examinations to be correlated, whether it is for predictive or for concurrent validity, should be of similar nature as also of similar objectives.

Certain assumptions, which have been made to pursue the present study, are as follows

1. The objective of the validity study is taken to be to see how well a student would do as a professional in his life.

2. In order to be able to study the criterion-related validity of any examination vis-a-vis the objective stated in (1), it is important to have a point of view on the type of education that is needed to train a citizen for tomorrow.

In this systems age, students must be increasingly trained to participate as contributors in interdisciplinary and analytical problem solving efforts, demanding from them multifaceted understanding of issues of technology and sciences. Thus, the knowledge of the basics of analytical subjects like Mathematics and Physics ~~is~~ maybe considered to be central to the student training in any discipline for preparing him to participate meaningfully in any profession-based social action. It is against this framework that, for the purpose of the validity study, this chapter assumes, amongst other examinations like Higher Secondary, the performances in subjects like concepts in science or the Modern Physics as appropriate candidates as predictors for the criterion-related validity analysis.

3. Further, for the purpose of the validity analysis under consideration, consistent with the above framework the criteria chosen are :

(i) Modern Physics and the Concepts in Science course performances,

(ii) Performances in several discipline courses ~~which is~~ given represented by the Grade Point Average (GPA's) for the disciplines, and

(iii) Students' overall performance which is represented by his latest Cumulative Grade Point Average (CGPA).

Understandably, the choice of these performances as criteria is based on the point of view that they (i.e. performances) have emerged from the details of the educational system, such as at BITS, which is essentially based on the theme of the interdisciplinary and broad-based pattern of professional education.

4. Finally, it is assumed that the same ability is being measured by various examinations. In fact, this has been established in chapter 2, where it has been found that examinations, in general, test Lower order skills of knowledge, comprehension and application

In addition to the criterion-related validity, this chapter also studies the theme of the concurrent validity. Need for such a study arises as a result of questions like - 'how far the performance in one discipline courses relates to the GPA of some other discipline courses'. The correlation thus obtained between the two concurrent measures would give the concurrent validity for the former performance measure ~~of examinations~~. It may be mentioned that the adjective "concurrent" is taken to mean examinations or assessments certainly within the Institute year and normally close to each other in terms of time-difference.

An important objective of the present study, in addition to studying the validity of examinations, is also to explore the best predictor and the best criterion under certain given situations.

4.5 Procedure of Estimating Criterion-related Validity

4.5.1 Predictive Validity

First of all the examination, on the basis of which the future performance can be predicted, is selected.

The chosen examination is called 'predictor'. Next, some other examination which matches in educational objectives with the requirements of the future performance under consideration ~~future course/examination~~ is selected. This is known as 'criterion' examination. Correlation between the marks in 'predictor' and 'criterion' examinations gives the measure of predictive validity* of the 'predictor' examination.

4.5.2 Concurrent Validity

Concurrent validity is nothing but the correlation between two contemporary examinations scores.

Both predictive and concurrent validity are represented by coefficient of correlations. Hence, validity is nothing but a coefficient of correlation with a definite connotation, as explained above, attached to it.

The interpretation of the coefficient of validity in the sections to follow would be done on the basis of following scale (Guilford, 1973):

TABLE 4.4
Interpretation of coefficient of correlation

Range of coefficient of correlation	Interpretation
Less than 0.20	Slight; almost negligible relationship
0.21 to 0.40	Low correlation; definite but small relationship
0.41 to 0.70	Moderate correlation; substantial relationship
0.71 to 0.90	High correlation; marked relationship
0.91 to 1.00	Very high correlation; very dependable relationship.

* From here onwards, for the purpose of the continuity of the running text, the term 'criterion-related validity' in terms of the correlation between the predictor-criterion pair may sometimes be used interchangeably with the term predictive validity.

4.6 Sample and Data Collection

To study the criterion related validity of the examinations under consideration following samples have been selected:

1. First sample consists of 310 students who were admitted to BITS in the year 1976. Presently, most of these students are in the second semester of their fifth academic year. Performance of each student as per following details was traced out:

(i) Percentage (normalised) marks obtained in Higher Secondary examination,

(ii) Grade in course PHY A211 - Modern Physics,

(iii) Grade point averages (GPAs) in Biology, Chemistry Mathematics, Physics, Core Science and Core Engineering courses, and, for any other discipline outside sciences, the GPA in own discipline courses, and, finally,

(iv) Latest CGPA.

2. In the other sample of 316 students, who were admitted to BITS in the year 1977, following information is considered:

(i) Percentage (normalised) marks obtained in Higher Secondary (H.S.) examination,

(ii) Grade in course SCI A111 - Concepts in Science,

(iii) Grade point averages (GPAs) in Biology, Chemistry, Mathematics and Physics courses, and, finally,

(iv) Latest CGPA.

Currently, most of the students from this 1977 batch are in the second semester of their fourth academic year.

For both the batches of students, higher secondary performance of each student in terms of normalised percentage was collected from Admission's Office of this Institute. Rest of the information was obtained from the computer centre of the Institute which keeps record of the performance of students who are studying in BITS.

Thus, the data on the selected sample consists of following:

TABLE 4.5

Samples selected for the predictive validity study

Sample	No. of students	Examinations	Performance in courses	Performance in discipline courses	Latest performance
1976 Batch	310	Normalised Performance in Higher Secondary examination	Grade in PHY A211 Modern Physics	GPA in BIO, CHEM, MATH, PHY, SCI, ENGG, CE, CHE, EEE, & ME Course	Latest CGPA
1977 Batch	316	Normalised Performance in Higher Secondary examination	Grade in SCI A111 Concepts in Science	GPA in BIO, CHEM, MATH, and PHY courses	Latest CGPA

For the purpose of convenience, various abbreviations used in this chapter are mentioned in ~~the~~ table 4.6 given on the next page:

TABLE 4.6

Abbreviations used in the present chapter

Full Name	Abbreviation
Higher Secondary	H.S.
Modern Physics	Mod Phy
Concepts in Science	Conc. SCI
Grade Point Average	GPA
Cumulative Grade Point Average	CGPA
Biology	BIO
Chemistry	CHEM
Physics	PHY
Mathematics	MATH
Science	SCI
Engineering	ENGG
Civil Engineering	CE
Computer Science	CS
Electrical & Electronics Engineering	EEE
Mechanical Engineering	ME

4.7 Results for the Criterion-related Validity Study

This chapter, in section 4.7.1, gives the results for predictive validity study, while, section 4.7.2, presents the results in terms of the concurrent validity study.

Needless to say, these results have extensively made use of the computer. The correlation programme, used here along with the other programmes used elsewhere in this thesis, is given in Appendix I.

4.7.1 The results of the predictive validity study are given through tables 4.7 to 4.21. Each table is self explanatory in terms of the information base.

TABLE 4.7

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for All the Students:

Sample	Criterion / Predictor	Modern Physics	Concepts in Science	GPA: BIO	GPA: CHEM	GPA: MATH	GPA: PHY	GPA: SCI	CGPA
1976 Batch 310 students	Higher Secondary	0.21		0.16	0.11	0.17	0.15	0.12	0.20
	Modern Physics			0.47	0.64	0.72	0.86	0.47	0.77
1977 Batch 316 students	Higher Secondary		0.24	0.10	0.18	0.23	0.41		0.20
	Concepts in Science			0.63	0	-0.52	-0.20		-0.09

TABLE 4.8

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Biology Students :

Sample	Criterion / Predictor	Modern Physics	Concepts in Science	GPA BIO	GPA CHEM	GPA MATH	GPA PHY	GPA SCI	CGPA
1976 Batch 11 Students	Higher Secondary	0.58		0.29	0.20	-0.04	0.66	-0.04	0.50
	Modern Physics			0.75	0.43	0.10	0.77	-0.23	0.75
1977 Batch 14 Students	Higher Secondary		-0.24	0.10	0.18	0.23	0.41		0.20
	Concepts in Science			0.63	0	-0.52	-0.20		-0.09

TABLE 4.9

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Chemistry Students:

Sample	Predictor \ Criterion	Modern Physics	Concepts in Science	GPA: BIO	GPA: CHEM	GPA: MATH	GPA: PHY	GPA: SCI	CGPA
1976 Batch 14 Students	Higher Secondary	0.44		0	0.08	-0.16	0.04	0.04	0.06
	Modern Physics			0.28	0.67	0.63	0.82	0.18	0.75
1977 Batch 34 Students	Higher Secondary		0.18	-0.02	0.15	0.17	0		0.01
	Concepts in Science			-0.29	-0.28	-0.20	-0.11		-0.08

TABLE 4.10

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Mathematics Students:

Sample	Predictor \ Criterion	Modern Physics	Concepts in Science	GPA BIO	GPA CHEM	GPA MATH	GPA PHY	GPA SCI	CGPA
1976 Batch 7 Students	Higher Secondary	-0.33		0.27	-0.42	-0.56	-0.47	-0.74	-0.41
	Modern Physics			0.38	0.88	0.81	0.98	0.71	0.94
	Higher Secondary		-0.13	0.10	-0.17	0.28	0.22		0.14
	Concepts in Science			0.05	0.42	0	0.15		0.20

TABLE 4.11

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Physics Students:

Sample	Criterion / Predictor	Modern Physics	Concepts in Science	GPA: BIO	GPA: CHEM	GPA: MATH	GPA: PHY	GPA: SCI	CGPA
1976 Batch 15 Students	Higher Secondary	-0.19		0.02	-0.04	-0.05	-0.19	-0.26	-0.24
	Modern Physics			0.18	0.69	0.77	0.84	0.67	0.84
1977 Batch 19 Students	Higher Secondary		0.37	0.19	0.22	-0.14	0.02		0.38
	Concepts in Science			0.28	-0.27	0.14	-0.19		0.95

TABLE 4.12

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Civil Engineering Students:

Sample	Criterion / Predictor	Mod Phy	Conc. SCI	GPA: BIO	GPA: CHEM	GPA: MATH	GPA: PHY	GPA: SCI	GPA: ENGG	GPA: CE	CGPA
1976 Batch 16 Students	Higher Secondary	0.76		0.52	0.14	0.39	0.56	0.07	0.56	0.39	0.60
	Modern Physics			0.46	0.28	0.46	0.69	0.10	0.39	0.31	0.51
1977 Batch 19 Students	Higher Secondary		0.24	0.37	-0.24	-0.01	-0.02				0.09
	Concepts in Science			-0.22	0.04	0.01	0.28				0.41

TABLE 4.13

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Chemical Engineering Students:

Sample	Criterion / Predictor	Mod Phy	Conc. SCI	GPA BIO	CHEM	MATH	PHY	SCI	ENGG	CHE	CGPA
1976 Batch 33 Students	Higher Secondary	0.01		0.10	-0.11	0.02	-0.03	0.03	-0.06	0.03	-0.01
	Modern Physics			0.49	0.17	0.08	0.25	0.64	0.14	0.32	0.50
1977 Batch 30 Students	Higher Secondary		0.23	0.10	-0.02	0.07	-0.14				0.20
	Concepts in Science			0.07	-0.06	-0.22	-0.06				0.26

TABLE 4.14

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Computer Science Students:

Sample	Criterion / Predictor	Mod Phy	Conc. SCI	GPA							
				BIO	CHEM	MATH	PHY	SCI	ENGG	CS	CGPA
1976 Batch 9 Students	Higher Secondary	0.22		0.13	-0.22	0.18	-0.12	0.25	-0.12	0.34	-0.17
	Modern Physics			0.26	0.45	0.70	0.80	0.63	-0.09	-0.28	0.45
1977 Batch 11 Students	Higher Secondary		0.29	0.38	0.48	0.34	0.72				0.72
	Concepts in Science			0.55	0.06	0.18	0.35				0.52

TABLE 4.15

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for EEE Students:

Sample	Criterion / Predictor	Mod Phy	Conc. SCI	GPA BIO	GPA CHEM	GPA MATH	GPA PHY	GPA SCI	GPA ENGG	GPA EEE	CGPA
1976 Batch 101 Students	Higher Secondary	0.12		0	-0.05	0.14	0.04	-0.09	0.05	0.06	0.04
	Modern Physics			0.26	0.47	0.58	0.77	0.32	0.45	0.57	0.70
1977 Batch 73 Students	Higher Secondary		0.16	-0.13	0.17	0.29					0.19
	Concepts in Science			0.21	0.32	0.11	0.38				0.42

TABLE 4.16

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Mechanical Engineering Students:

Sample	Criterion / Predictor	Mod Phy	Conc. SCI	GPA							CGPA
				BIO	CHEM	MATH	PHY	SCI	ENGG	ME	
1976 Batch 56 Students	Higher Secondary	0.03		0.20	0	-0.02	-0.02	0.22	0.09	0.15	0.34
	Modern Physics			-0.10	0.10	0.31	0.63	0.15	0.04	0	0.36
1977 Batch 45 Students	Higher Secondary		0.02	-0.06	-0.28	-0.23	-0.27				-0.30
	Concepts in Science			0.25	-0.02	0.28	0.30				0.31

TABLE 4.17

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Instrumentation Students:

Sample	Criterion / Predictor	Modern Physics	Concepts in Science	GPA					CGPA
				BIO	CHEM	MATH	PHY	SCI	
1976 Batch 11 Students	Higher Secondary	0.14		0.01	-0.18	0.47	0.52	0.14	0.46
	Modern Physics			0	-0.14	0.77	0.63	0.39	0.63
1977 Batch 6 Students	Higher Secondary		0.42	-0.29	-0.18	0.34	-0.38		0.10
	Concepts in Science			-0.26	-0.51	-0.28	-0.73		-0.44

TABLE 4.18

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Pharmacy Students:

Sample	Criterion / Predictor	Modern Physics	Concepts in Science	GPA BIO	GPA CHEM	GPA MATH	GPA PHY	GPA SCI	CGPA
1976 Batch 7 Students	Higher Secondary	-0.37		0.04	-0.14	0.04	-0.07	0.42	-0.32
	Modern Physics			0.38	0.26	-0.09	0.62	0	0.79
1977 Batch 3 Students	Higher Secondary		0	-0.55	-0.37	-0.83	-0.71		-0.63
	Concepts in Science			0	0	0	0		0

TABLE 4.19

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Management Students:

Sample	Criterion / predictor	Modern Physics	Concepts in Science	GPA					CGPA
				BIO	CHEM	MATH	PHY	SCI	
1976 Batch 29 Students	Higher Secondary	0.48		0.23	0.22	0.16	0.23	0.12	0.01
	Modern Physics	–		0.36	0.33	0.26	0.53	0.62	0.37
1977 Batch 30 Students	Higher Secondary		-0.10	-0.06	0.25	0.45	0.14		0.15
	Concepts in Science			0.73	0.86	0.51	0.77		0.13

TABLE 4.20

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for Economics Students:

Sample	Criterion / Predictor	Concepts in Science	GPA				CGPA
			BIO	CHEM	MATH	PHY	
1977 Batch 13 Students	Higher Secondary	0.34	0.49	0	-0.52	-0.59	-0.19
	Concepts in Science		0.04	0.61	-0.01	0.18	0.51

TABLE 4.21

Correlation Coefficients between the Performances in Predictor and Criterion Examinations for English Students:

Sample	Criterion / Predictor	Concepts in Science	GPA BIO	GPA CHEM	GPA MATH	GPA PHY	CGPA
1977 Batch 4 Students	Higher Secondary	0.76	0	0.19	-0.01	0	0.76
	Concepts in Science		0	0.77	-0.55	0.44	0.80

4.7.2 Results for Concurrent Validity Study

As mentioned earlier, this section reports the results on the concurrent validity study which, consistent with the constraint of the availability of the adequate data, has been carried out on the 1976 sample only. The results are presented in table 4.22 which is self-explanatory.

TABLE 4.22

Correlation Coefficients between the Performances in Various Concurrent Examinations for Civil, Chemical, Electrical & Electronics and Mechanical Engineering Students:

Samples 1976 Batch	Examination / Examination	CE	CHE	EEE	ME	CGPA
16 CE Students	ENGG	0.89				0.81
33 CHE Students	ENGG		0.87			0.53
101 EEE Students	ENGG			0.88		0.79
56 ME Students	ENGG				-0.09	-0.03

4.8 Analysis of Results on Criterion-Related Validity Study

4.8.1 Assumptions

The analysis proceeds with the following assumptions:

(i) 1976 input of 310 students and 1977 input of 316 students are taken to be representative samples. Thus the samples are taken to be easily substitutable for each other in terms of arriving at general conclusions.

(ii) Validity analysis is conducted for a spectrum of disciplines belonging to groups ~~consisting of~~:

(a) Science disciplines : BIO, CHEM, PHY and MATHS ;

(b) Engineering disciplines : CE, CHEM, EEE and ME ;

(c) Applied Science and technique oriented disciplines : CS, INSTR, PHA

and (d) Social Science and Humanities disciplines : MGTS, ECON* and ENGL*

(iii) For the purpose of the analysis, following spectrum of criteria is considered: BIO GPA, CHEM GPA, MATH GPA, PHY GPA, SCI GPA+ and CGPA.

4.8.2 Analysis for the 1976 Batch for the choice of the Best Predictor

This section gives the analysis for the 1976 batch for the choice of the best predictor amongst performances in the higher secondary examinations and the examination

---

*Only for 1977 batch.

+Only for 1976 batch

for the modern physics course. The comparison of the performances in terms of the above mentioned predictors for the representative samples of the students at the Institute as generated through the batch of 1976 input is shown in figures 4.1 and 4.2. If the shaded portions of the bars presented in figures 4.1 and 4.2 describe the above stated comparison in terms of the percentage frequency of the event for a given predictor, when viewed against the entire spectrum of criteria from GPAS in biology, chemistry mathematics, physics science to CGPA, giving correlations between levels moderate to very high; the dotted bars present the comparison in terms of the frequency of the event for the predictor under consideration, once again when viewed against the total spectrum of the above stated criteria, giving correlations ranging from low to negative.

As regards to figure 4.3, for the entire sample of 310 students from the 1976 batch, it gives percentage correlation frequencies corresponding to each of the levels from negative to very high for both the predictors, when viewed against each of the criteria under consideration.

To begin with, apart from investigating the validity of a given predictor, the study applies itself to the question as to which is the best predictor for the samples under study, for the entire spectrum of the proposed criteria. Towards this, various sub-samples as available from 1976 batch are considered first.

The analysis as emerging from the results, recorded through in figures 4.1 to 4.3, then, is as follows:

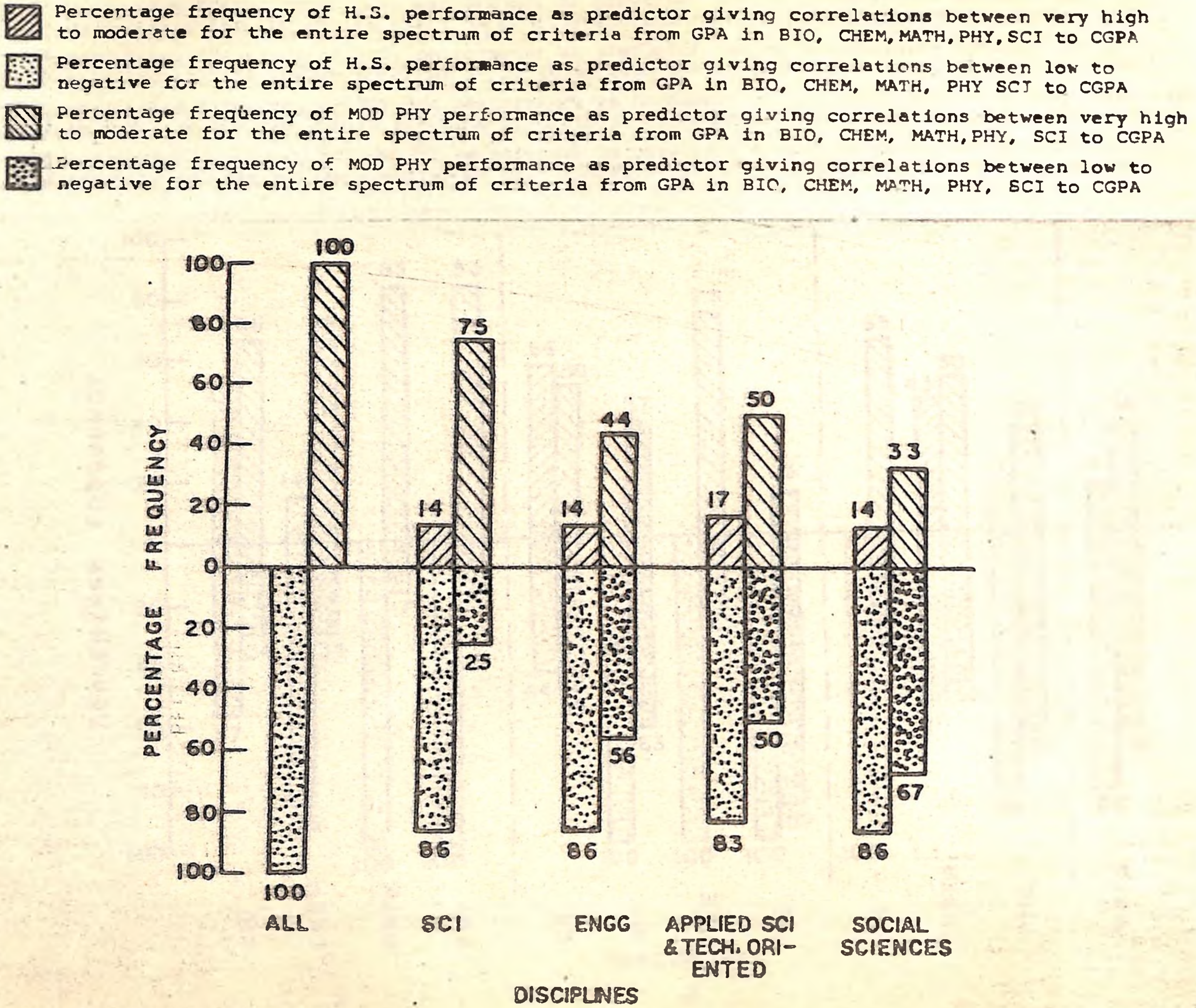


FIG. 4.1 COMPARISON BETWEEN THE PERFORMANCES IN HIGHER SECONDARY AND MODERN PHYSICS AS PREDICTOR FOR VARIOUS GROUPS OF DISCIPLINES FROM 1976 BATCH.

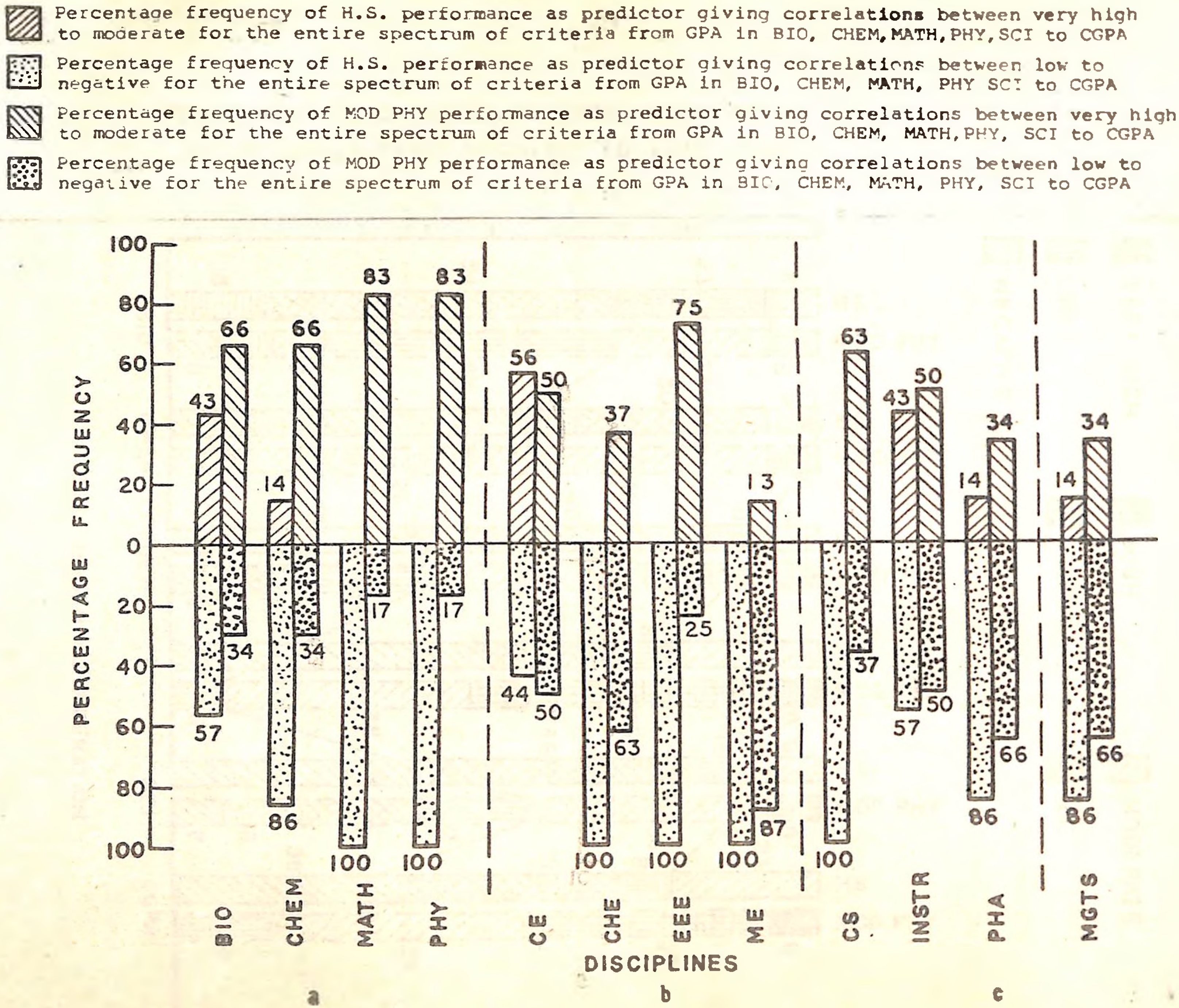


FIG. 4.2 COMPARISON BETWEEN THE PERFORMANCES IN HIGHER SECONDARY AND MODERN PHYSICS AS PREDICTOR FOR VARIOUS DISCIPLINES FROM 1976 BATCH.

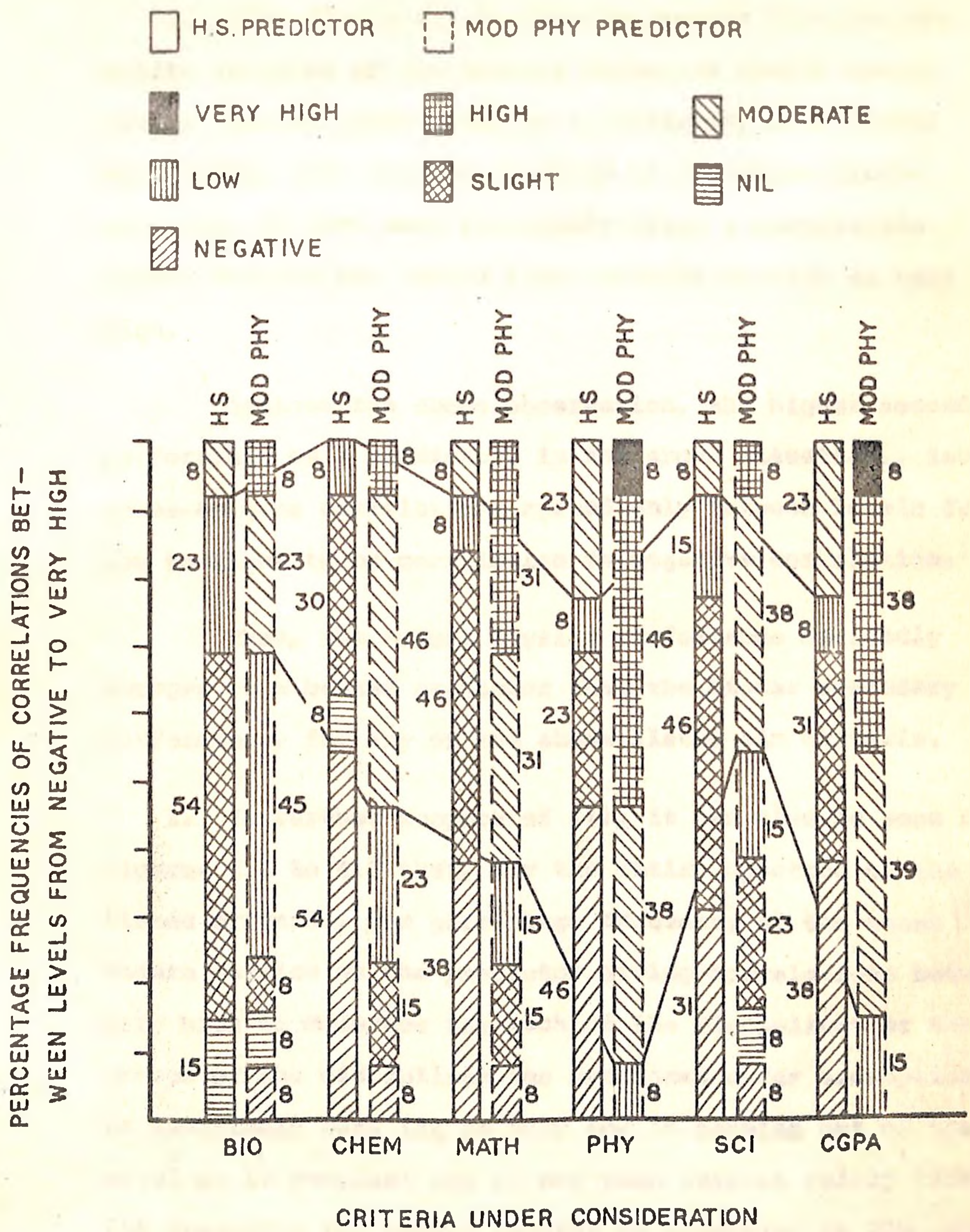


FIG. 4.3 COMPARISON OF PERFORMANCES IN HIGHER-SECONDARY AND MODERN PHYSICS EXAMINATIONS AS PREDICTORS, WHEN VIEWED AGAINST ENTIRE SPECTRUM OF THE CRITERIA UNDER CONSIDERATION, FOR 1976 BATCH.

1. From figure 4.1 it clearly emerges that, for the entire universe of the student number as sample space, Modern Physics performance as a predictor, when viewed against the wide ranging spectrum of the above listed criteria, ~~in each case~~ invariably gives a correlation spread between the levels from moderate to high to very high.

Against the above observation, the higher secondary performance as a predictor, in the above situations, invariably ~~seems to give~~ reports correlations spread only between levels from low to slight to no correlation to negative correlation.

Thus, the Modern Physics performance decidedly emerges as a better predictor than the higher secondary performance, for any of the above listed six criteria.

2. In further support of (1), it can also be seen from figures 4.1 to 4.2 that, for the entire spectrum of the above listed criteria, the percentage frequency of the event Modern Physics as the predictor giving correlations between very high to moderate for each of the disciplines or the groups of the disciplines (as mentioned under assumptions) is invariably more (it is more for 16 samples out of the total of 17 samples) and if not then atleast fairly close (if frequency for Modern Physics as predictor is 50%, the same for Higher Secondary is 56%) to the corresponding frequency of the event relating to the higher secondary performance as the predictor.

Similarly, from the above figures, it also follows that the percentage frequency of the event Modern Physics performance as a predictor giving correlations between low to negative for the entire spectrum of disciplines or their groups is invariably less and if not then atleast fairly close to the corresponding percentage frequency of the event relating to the Higher Secondary performance as the predictor.

3. On a critical analysis of the figure 4.1, it can be seen that while, for the science category, the percentage frequency of the Modern Physics performance as a predictor giving correlations between very high to moderate (is of the order of 75%), the corresponding percentage frequencies for the categories of Engineering, Applied Sciences & Technique Oriented disciplines and Social Sciences are of the order of 44%, 50% and 33%, respectively.

Thus, it emerges that, while Modern Physics performance is a better predictor than the Higher Secondary performance, within the various groups of disciplines the Modern Physics performance is more suitable as a predictor for science group of students than for the students from the Applied Science and Technique Oriented disciplines than for Engineering group of stude than for Management group of students, in that order.

4. From figure 4.2(a) it can be further seen that, amongst the Science group, Modern Physics is more suitable as a predictor for the disciplines of Mathematics and Physics than for the disciplines of Biology and Chemistry.

Similarly, from 4.2(b), it can be seen that, amongst the Engineering group, Modern Physics is more suitable as a predictor for the discipline of EEE than for the discipline of Civil Engineering than for the discipline of Chemical Engineering, than for the discipline of Mechanical Engineering, in that order.

From figure 4.2 (c), it can be seen that amongst the Applied Science & Technique Oriented disciplines, the Modern Physics performance is more suitable as a predictor for the discipline of computer science than for the discipline of Instrumentation, than for Pharmacy, in that order.

5. Finally, coming to figure 4.3, it can be seen that it further comprehensively summarises the observation stated under ~~the~~ point (1) of this section. The figure 4.3, consistent with the observations from figure 4.1, clearly shows that for entire sample of the 1976 batch, the performance in Modern Physics course as predictor, for each of the criteria under consideration, invariably gives percentage frequency of correlations ranging from levels moderate to very high, much higher than that for the Higher Secondary performance as predictor; thus implying that Modern Physics is a superior predictor to Higher Secondary examination. In quantitative terms, if the percentage frequencies in the above context for the Modern Physics as the predictor vary from 31% to 92%, the same for the Higher Secondary vary from 0% to 23%.

With this, now, one can proceed with the analysis for the 1977 batch for the choice of the best predictor.

4.8.3 Analysis for the 1977 Batch for the Choice of the Best Predictor

This section gives the analysis for the 1977 batch in terms of the choice of the best predictor amongst performances in the Higher Secondary examination and the examination for the Concepts in Science course. As already reported under assumption (ii) of section 4.8.1, this sample, under the category of Social Sciences & Humanities, in addition to the discipline of Management, also includes the disciplines of Economics and English, and thus covers a little broader spectrum of the Universe ~~samples~~ than in the case of sample from the 1976 batch. However, as regards to the spectrum of the criteria, the analysis of 1977 batch covers only 5 criteria; namely, BIO GPA, CHEM GPA, MATH GPA, PHY GPA and CGPA; thus excluding the SCI GPA unlike in the case of the analysis for the 1976 batch.

It is then, within the above frame work, that below is presented the validity analysis for 1977 batch under consideration.

1. Comparision of the performances in the Higher Secondary examination and the Concepts in Science course examination for the representative sample of the students at the institute as reflected through the 1977 batch is given in figures 4.4 to 4.6. While the shaded portions of the bars presented describe the above stated comparison in terms of the percentage frequency for given event for a predictor, when viewed against the entire spectrum of criteria from BIO GPA, CHEM GPA, MATH GPA, PHY GPA to CGPA, giving correlations between levels moderate to very high; the dotted bars present the comparison in terms of the percentage frequency of the event for the predictor under consideration, once again when viewed against the total spectrum of the above stated criteria, giving correlations ranging from low to negative.

It is through a critical analysis of the above stated figure 4.4, it emerges that, for the entire universe of the student number as the sample space, Higher Secondary as a predictor, when viewed against the entire spectrum of the above mentioned criteria, has the percentage frequency for the event corresponding to the correlations being between moderate to very high only 20%, while the percentage frequency for the correlations being between levels low to negative is as high as 80%.

2. Second predictor under consideration for 1977 batch is the performance in Concepts in Science course. From figure 4.4 (a), it clearly emerges that, for the entire

Percentage frequency of H.S. performance as predictor giving correlations between very high to moderate for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of H.S. performance as predictor giving correlations between low to negative for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of CONC. SCI performance as predictor giving correlations between very high to moderate for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of CONC. SCI performance as predictor giving correlations between low to negative for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

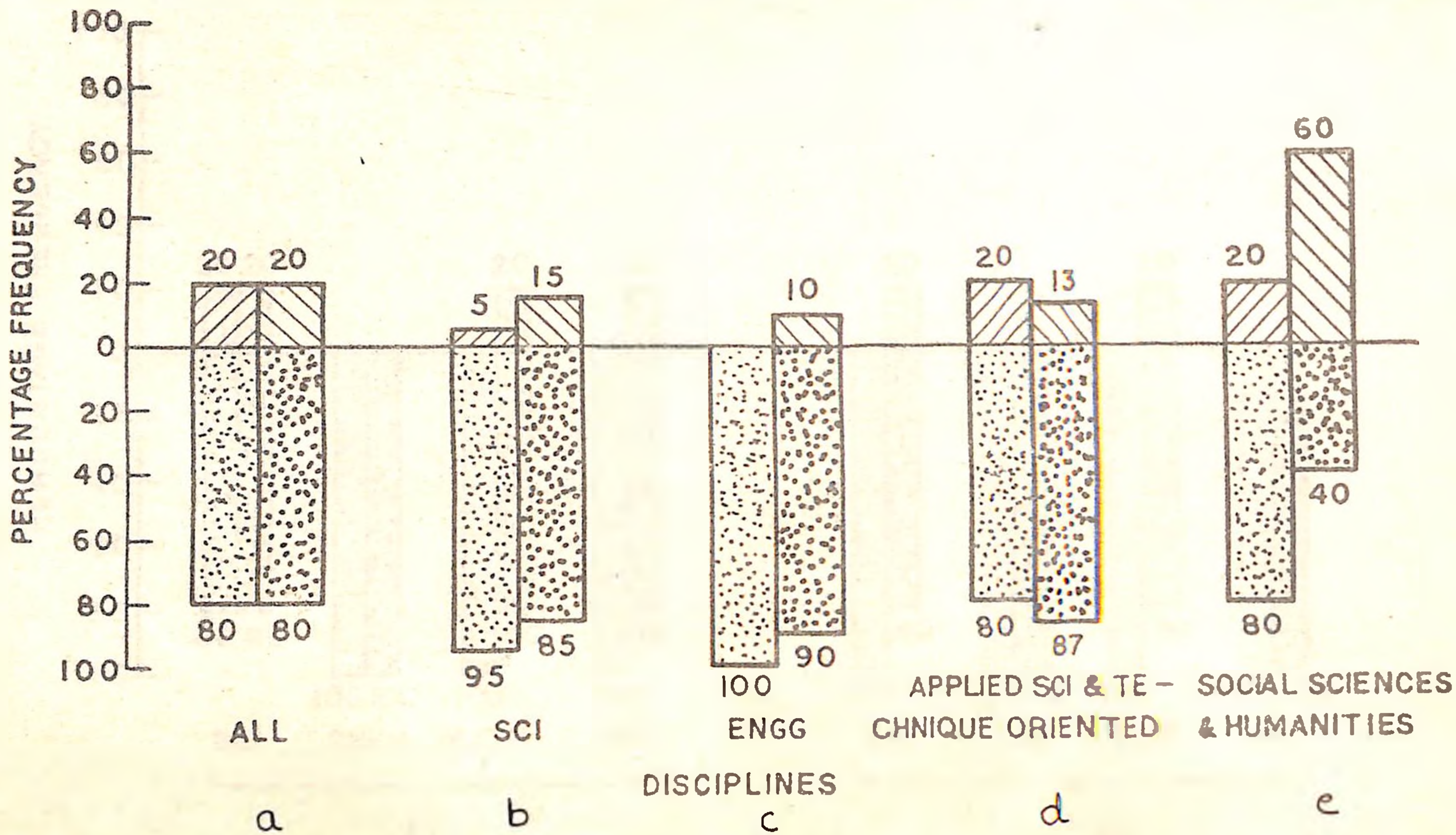


FIG. 4.4 COMPARISON BETWEEN THE PERFORMANCES IN HIGHER SECONDARY AND CONCEPTS IN SCIENCE AS PREDICTOR FOR VARIOUS GROUPS OF DISCIPLINES FROM 1977 BATCH.

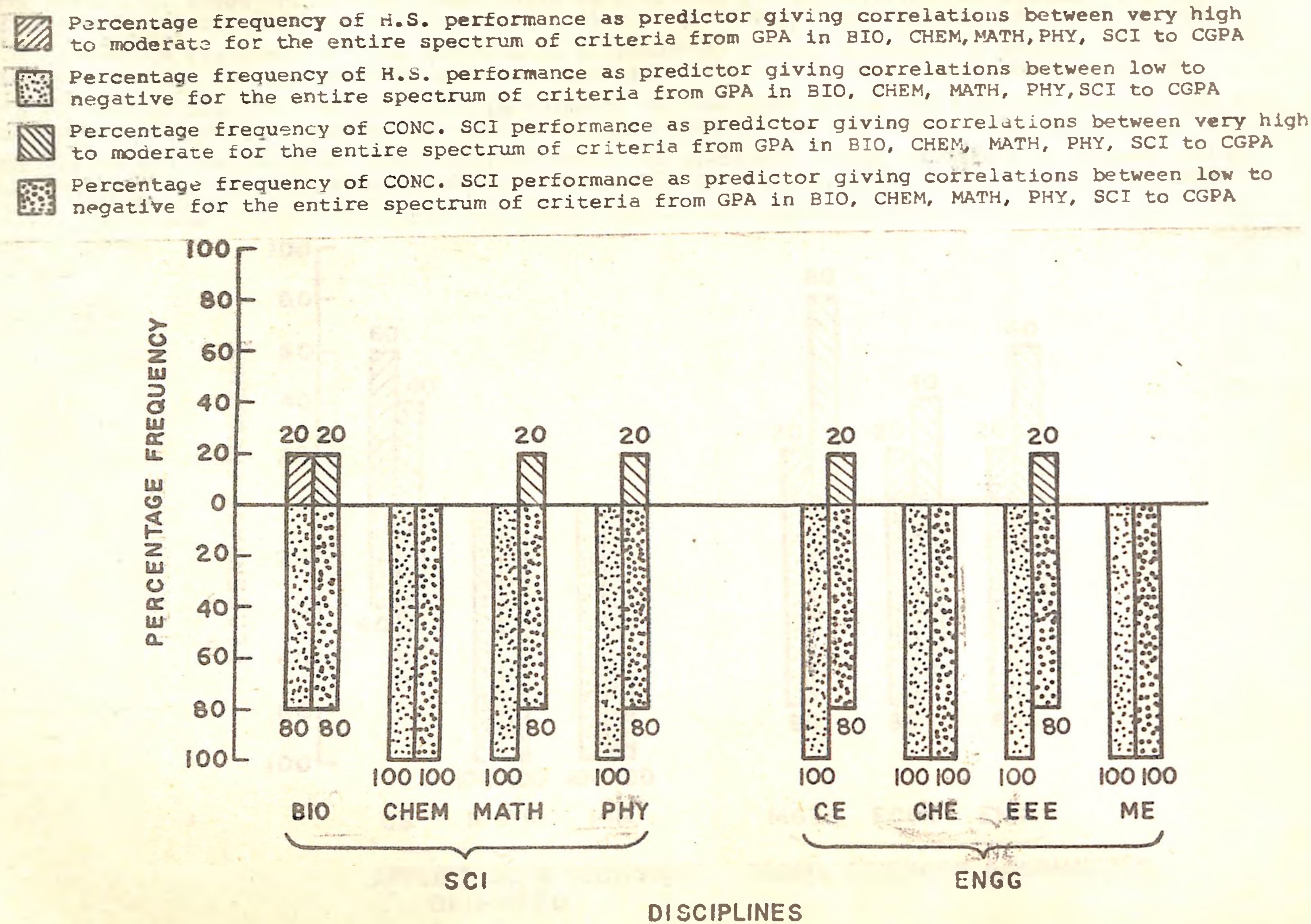


FIG. 4.5 COMPARISON BETWEEN THE PERFORMANCES IN HIGHER SECONDARY AND CONCEPTS IN SCIENCE AS PREDICTOR FOR VARIOUS DISCIPLINES FROM 1977 BATCH.

Percentage frequency of H.S. performance as predictor giving correlations between very high to moderate for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of H.S. performance as predictor giving correlations between low to negative for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of CONC. SCI performance as predictor giving correlations between very high to moderate for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

Percentage frequency of CONC. SCI performance as predictor giving correlations between low to negative for the entire spectrum of criteria from GPA in BIO, CHEM, MATH, PHY, SCI to CGPA

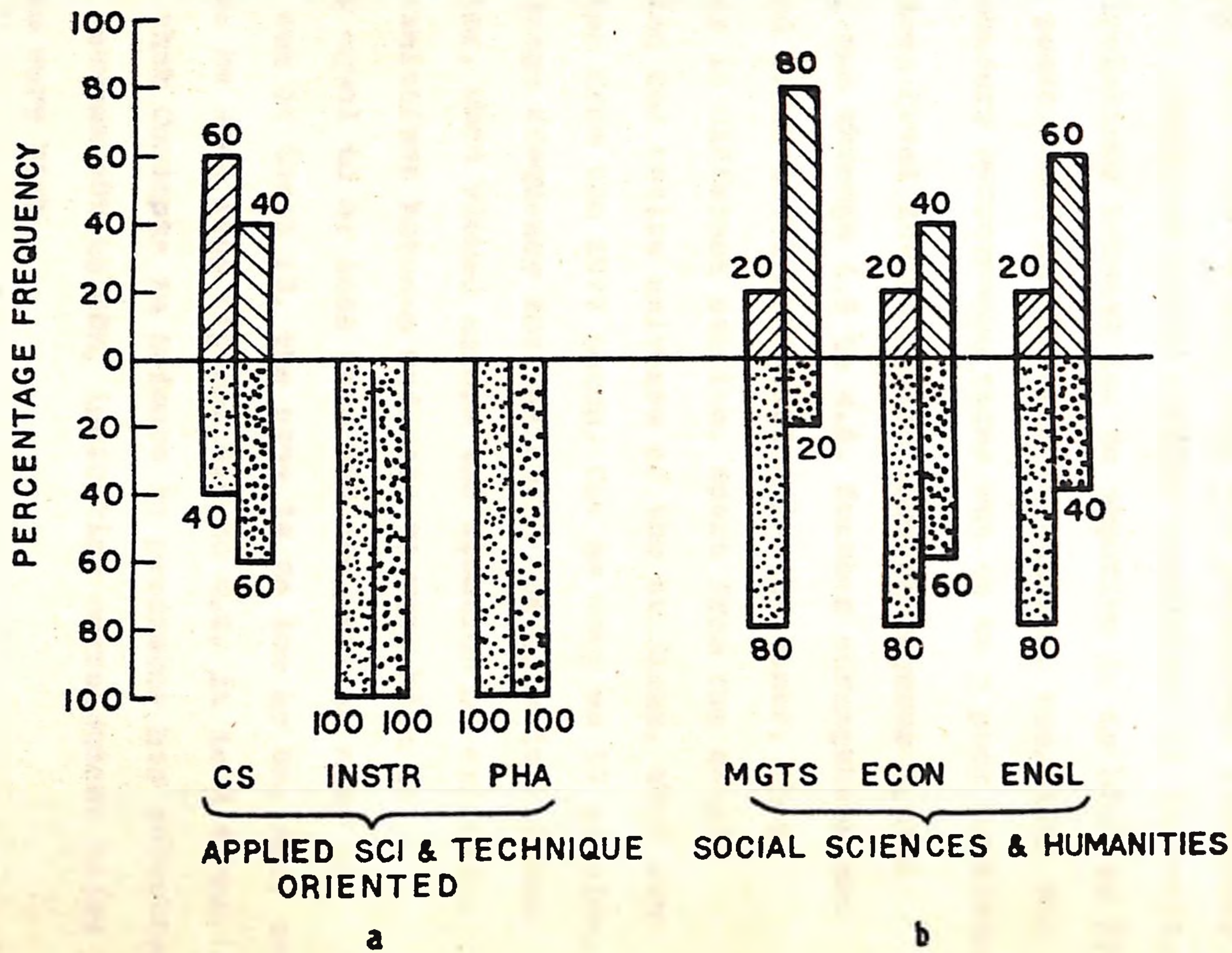


FIG. 4.6 COMPARISON BETWEEN THE PERFORMANCES IN HIGHER SECONDARY AND CONCEPTS IN SCIENCE

sample of the 1977 students, while the percentage frequency of the event Concepts in Science as predictor, when viewed against the range of criteria, giving correlation between moderate to very high is only 20%, the percentage frequency for the same predictor under similar conditions of the criteria giving correlations between low to negative is as high as 80%. Hence the performance in Concepts in Science, too, like the Higher Secondary performance turns out to be a poor predictor.

2. A micro level investigation through figures 4.4(b) to 4.4(e), and through 4.5 to 4.6, further strengthens the above broad macro level observation in the sense, when checked for 18 different samples, apart from the sample constituting the entire universe of the students, that can be generated from the 1977 batch, for as many as 13 samples, the percentage frequency for the event Concepts in Science as predictor, when viewed across the spectrum of criteria, giving correlations between moderate to very high is invariably equal to or less than 20% and, for as many as 5 samples out of these 13, the same is as low as 0%. And, as can further be seen from figures 4.4 to 4.6, it is 5 times out of 18 that Concepts in Science as predictor has percentage frequency between 40% to 60%, indicating correlations being moderate to very high.

Against this, the Higher Secondary performance only once has its corresponding percentage frequency as high as 60%, for the other remaining 17 samples, the same being invariably equal to or lower than 20%.

Thus, concepts in science, though a poor predictor by itself, may be considered to be a slightly better predictor than the Higher Secondary performance.

3. On a further critical analysis of the figure 4.4, it emerges that the Concepts in Science ~~seems to be~~ is a good predictor only for the Social Sciences and the Humanities disciplines.

4. Thus, putting together the observations for the previous sub-section and this sub-section and assuming a temporal invariance in the character of the two batches of 1976 and 1977, one can conclude that on the whole, whatever may, normally, be the criterion, the performance in Modern Physics is a far better predictor than the performance in Concepts in Science as also the Higher Secondary performance; and, if forced to discriminate further, then that the performance in Concepts in Science, though by itself a poor predictor, is a slightly better predictor than the performance in Higher Secondary. In such case then, one can also conclude that normally, the performance in the institute - based examination is a better predictor than the performance in the Higher Secondary examinations.

Indeed, the observation regarding Concepts in Science being a poor predictor warrants further query and can be pursued in the futuristic context. One possible reason for this could be the lack of content correlation between what has been taught in the course with what was to follow in the later years.

It is against the above frame of reference that the remaining part of the chapter restricts its attention to the further perusal of Modern Physics as the predictor in terms of the investigation on what can be considered to be the best criterion for it.

4.8.4 The Analysis of the Modern Physics as the Predictor for the choice of the Best Criterion

This section studies the question : what can be considered to be the best criterion if Modern Physics were to be the predictor ? Understandably, the study, for this purpose, exploits the data-base available from the 1976 batch.

In the analysis, samples from the individual science disciplines would be discussed first, followed by the samples from Engineering disciplines and those from Applied Science & Technique Oriented disciplines, in that order. Finally the study will analyse the entire 1976 batch taken together.

1. Science Discipline-wise Analysis

Science-discipline-wise

The correlation levels between the predictor as represented by the performance in the Modern Physics course and each of the criteria taken separately from the spectrum of criteria ranging from BIO GPA to MATH GPA to PHY GPA to CGPA are given in figure 4.7.

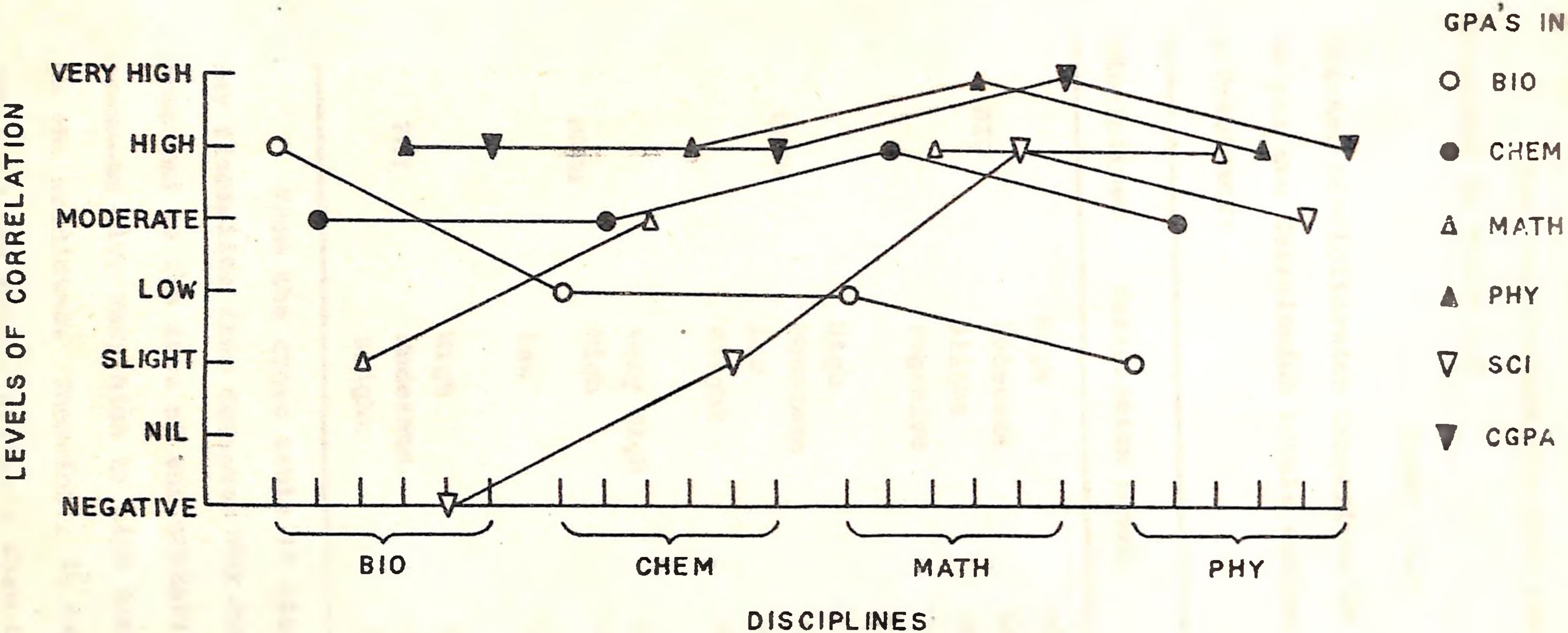


FIG. 4.7 SCIENCE DISCIPLINE-WISE CORRELATION LEVELS FOR EACH OF CRITERION AGAINST MODERN PHYSICS AS THE PREDICTOR.

The observations emerging from figure 4.7 are reported in table 4.8

TABLE 4.23

Science Disciplinewise Comparison of the various criteria as per the Correlation Levels against Modern Physics as a Predictor:

Disciplines	Correlation Level	Criterion
BIO	High	BIO GPA, PHY GPA and CGPA
	Moderate	CHEM GPA
	Slight	MATH GPA
	Negative	SCI GPA
CHEM	High	PHY GPA and CGPA
	Moderate	CHEM GPA and MATH GPA
	Low	BIO GPA
	Slight	SCI GPA
MATH	Very High	PHY GPA and CGPA
	High	CHEM GPA, MATH GPA and SCI GPA
	Low	BIO GPA
PHY	High	MATH GPA, PHY GPA and CGPA
	Moderate	CHEM GPA and SCI GPA
	Slight	BIO GPA

From the above table it clearly emerges that, for any discipline from Sciences PHY GPA/CGPA as criterion, when compared to the rest of the criteria under consideration, ~~seems to~~ gives very high to high correlation for Modern Physics as the predictor. Therefore, it can be concluded that PHY GPA and CGPA are better criteria than the others.

2. Engineering Disciplinewise Analysis

engineering-discipline-wise
The correlation levels between the predictor as represented by the performance in the Modern Physics course and each of the criteria, taken separately from the spectrum of criteria ranging from BIO GPA to CHEM GPA to MATH GPA to PHY GPA to SCI GPA to CGPA; are given in figure 4.8.

The observation emerging from figure 4.8 are reported in table 4.24.

TABLE 4.24.

Engineering Disciplinewise Comparison of the Various Criteria as per the Correlation Levels against Modern Physics as a Predictor:

Disciplines	Correlation Level	Criterion
CE	Moderate	BIO GPA, MATH GPA, PHY GPA & CGPA
	Low	CHEM GPA
	Slight	SCI GPA
CHE	Moderate	BIO GPA, SCI GPA and CGPA
	Low	PHY GPA
	Slight	CHEM GPA and MATH GPA
EEE	Very High	PHY GPA
	Moderate	CHEM GPA, MATH GPA and CGPA
	Low	BIO GPA and SCI GPA
ME	Moderate	PHY GPA
	Low	MATH GPA and CGPA
	Slight	CHEM GPA and SCI GPA
	Negative	BIO GPA

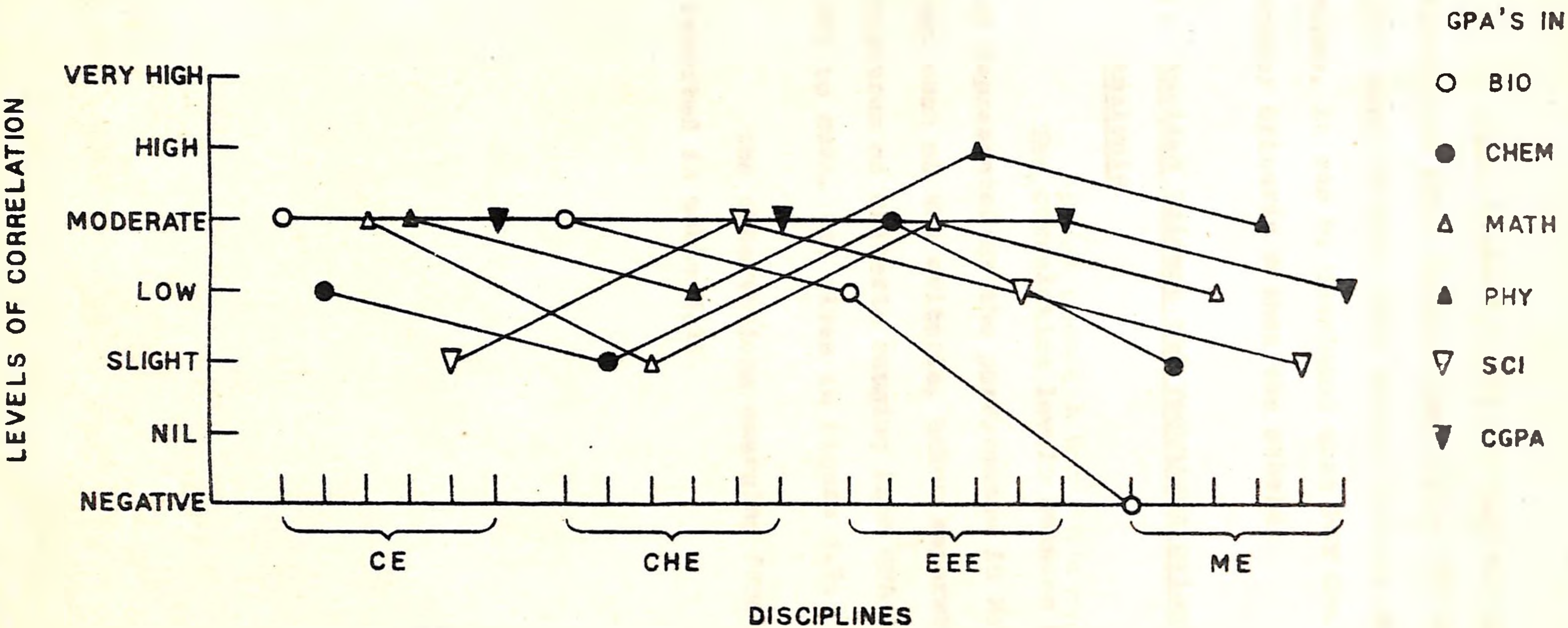


FIG. 4.8 ENGINEERING DISCIPLINE-WISE CORRELATION LEVELS FOR EACH OF THE CRITERION AGAINST MODERN PHYSICS AS THE PREDICTOR.

From table 4.24. it clearly emerges that, for any discipline from Engineering, either PHY GPA or CGPA give good correlations with Modern Physics as the predictor. Hence, it can be concluded that PHY GPA and CGPA are better criteria ~~a~~ than the others.

3. Applied Science and Technique Oriented Disciplinewise Analysis

The Applied science & Technique oriented-discipline-wise correlation levels between the predictor as represented by the performance in Modern Physics course and each of the criteria, taken separately, from the spectrum of criteria ranging from GPA in BIO to MATH to PHY to CGPA, are given in figure 4.9.

The observations emerging from figure 4.9 are reported in table 4.25.

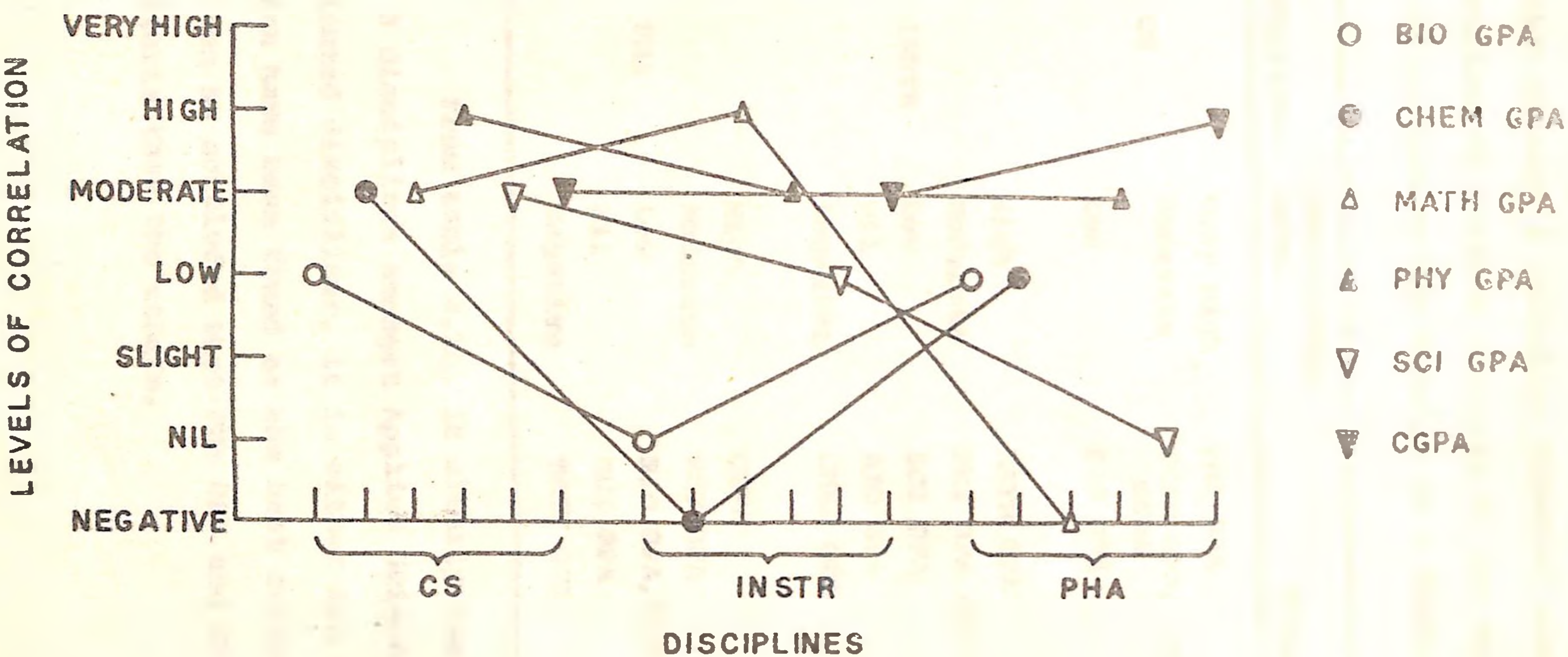


FIG. 4.9 APPLIED SCIENCE & TECHNIQUE ORIENTED DISCIPLINE-WISE CORRELATION LEVELS FOR EACH OF THE CRITERION AGAINST MODERN PHYSICS AS THE PREDICTOR.

TABLE 4.25

Applied Science & Technique Oriented Disciplinewise Comparison of various criteria as per the Correlation Levels Against Modern Physics as a Predictor.

Disciplines	Correlation Level	Criterion
CS	Very High	PHY GPA
	Moderate	CHEM GPA, MATH GPA, SCI GPA and CGPA
	Low	BIO GPA
INSTR	High	MATH GPA
	Moderate	PHY GPA and CGPA
	Low	SCI GPA
	Nil	BIO GPA
	Negative	CHEM GPA
PHA	High	CGPA
	Moderate	PHY GPA
	Low	BIO GPA, CHEM GPA
	Nil	SCI GPA
	Negative	MATH GPA

From table 4.25, it clearly emerges that in 2 out of 3 disciplines amongst Applied Science and Technique Oriented disciplines, it is either GPA PHY or the CGPA which have been found as the best criteria. Therefore, it can be concluded that PHY GPA and CGPA are better criteria than the others.

4. **Analysis for the Science Group of Disciplines, Engineering Group of Disciplines and Applied Science & Technique Oriented Group of Disciplines:**

For various groups of disciplines, the correlations between the Modern Physics as the predictor against each of the criteria, namely, BIO GPA, CHEM GPA, MATH GPA, PHY GPA, SCI GPA and CGPA, are shown in figure 4.10. More specifically ~~Further~~, the figure shows the percentage frequencies of the correlations being between levels moderate to very high for the above indicated group of disciplines, as against a single disciplinewise analysis as presented earlier. The groups of disciplines covered are Science group, Engineering group and Applied Science & Technique Oriented group.

The observations emerging from figure 4.10 are reported in table 4.26.

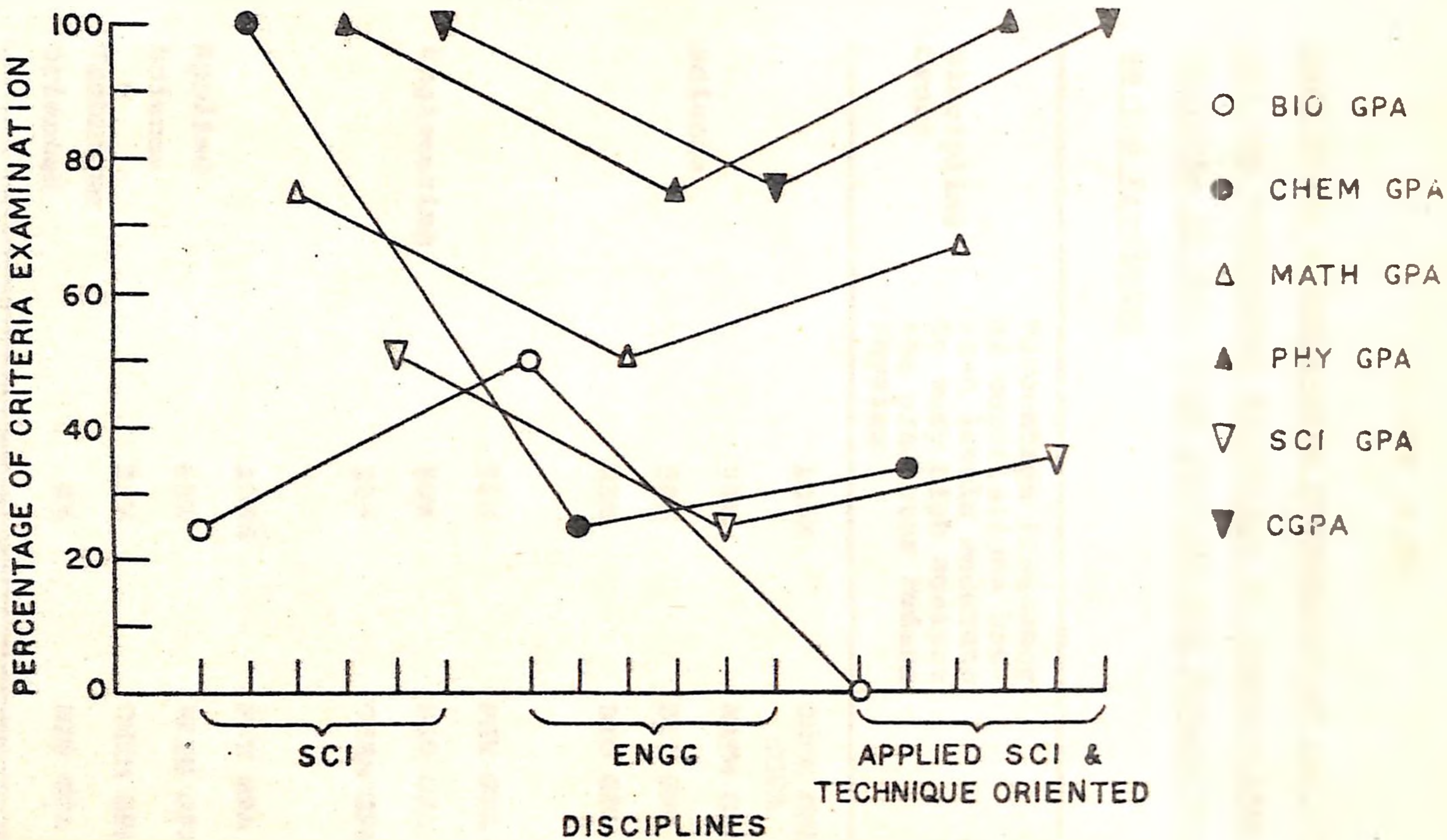


FIG. 4.10 DISCIPLINE - GROUP - WISE COMPARISON OF THE VARIOUS CRITERIA AS PER THE PERCENTAGE FREQUENCIES OF THE CORRELATIONS BETWEEN MODERATE TO VERY HIGH AGAINST MODERN PHYSICS AS A PREDICTOR.

TABLE 4.86.

Discipline - Group-wise Comparison of the Criterion, as per the percentage frequency of Correlations between Levels Moderate to Very High against the Modern Physics performance as the Predictor

Discipline Group	Percentage frequency of correlations between levels moderate to very high against the predictor Modern Physics	Criterion
Science	100%	CHEM GPA, PHY GPA and CGPA
	75%	MATH GPA
	50%	SCI GPA
	25%	BIO GPA
Engineering	75%	PHY GPA and CGPA
	50%	BIO GPA and MATH GPA
	25%	CHEM GPA and SCI GPA
Applied Science & Technique Oriented	100%	PHY GPA and CGPA
	65%	MATH GPA
	34%	CHEM GPA and SCI GPA
	0%	BIO GPA

Thus, as observed in subsections (1) to (3) of this section, PHY GPA and CGPA are invariably emerging as the good criteria against Modern Physics as the predictor for all the discipline groups described above.

An interesting side conclusion also seems to emerge from the above analysis. While it is observed that Modern Physics and PHY GPA/CGPA, as per the correlation level-wise analysis seem to make the best predictor-criteria pairs for any sample, it may be worthwhile to answer a question as to for which sample the above stated combination(s) are more suitable. Table 4.27 ~~tries to~~ presents the analysis for the same.

TABLE 4.27.

Discipline - Group-wise samples placed in the order of their percentage frequency showing correlations between moderate to very high for each of the Criterion-predictor pairs in terms of PHY GPA-Modern Physics and CGPA—Modern Physics respectively

Discipline Group	Percentage Frequency of correlation levels ranging from moderate to very high for	
	MOD PHY-PHY GPA pair	MOD PHY-CGPA pair
Science Group	100%	100%
Applied Science and Technique Oriented Group	100%	100%
Engineering Group	50%	75%

Thus, while, as observed through the subsections (1) to (4) of this section, PHY GPA and CGPA are emerging as the best criteria against the Modern Physics as predictor for each of the discipline or the groups of the disciplines, table 4.17 shows that the above predictor criteria pairs, while being the best choices possible for each of the disciplines or the groups of disciplines, are more suitable for the Science and the Applied Science & Technique Oriented groups than for the Engineering group of disciplines ~~in that order~~. And, further analysing these two pairs with respect to the Engineering group, it emerges that for this sample the Modern Physics - CGPA makes a little better pair than the Modern Physics - PHY GPA pair.

4.8.5 Analysis of the Results on the Concurrent Validity Study

To study the concurrent validity, correlations between two contemporary examinations or performances have been considered. Different concurrent sets selected for the study are - correlations between ENGG courses and student's own discipline courses (CE, CHE, EEE and ME) or ENGG courses and CGPA. Correlations emerging between various concurrent measures can be seen from table 4.22. For the reason of the availability of the appropriate data-base, the concurrent validity study has been persued in terms of the 1976 batch only.

Looking at table 4.22, most of the correlations between two concurrent examinations, except for the sample of ME students are found to be consistent in the sense that they vary only between moderate to very high levels.

Indeed, the details pertaining to the sample of ME students are rather inconsistent. Table 4.16 also in some sense provides a prelude for such inconsistency in the sense that even with Modern Physics as predictor against any of the criteria, except PHY GPA, for this sample of students, the predictive validity ranges only between levels negative to low; the same for PHY GPA as criterion being as high as moderate only. Another possible indication to expect such an inconsistent performance from this particular batch of ME students, seems to emerge from figure 4.2 (b), where even for Modern Physics as predictor, when viewed against the entire spectrum of the criteria, for this sample of ME student, the percentage frequency of correlations being between levels moderate to very high is only 17%. No doubt for a given section of small number of students, in such a manner, to give inconsistent indices of its performance pattern is a rather peculiar thing and must be a result of a complex combination of various factors contributed by the course contents, the teacher and the student; thus requiring an investigation clearly falling outside the scope of the present work. Thus, in other words, in a futuristic context, it may be mentioned that instead of generalising anything on immediate basis in terms of these observations, it will be helpful to first, further, in a detailed and structured manner investigate the test reliability, the content validity as also the content correlations specifically for the various courses studied by only this group of students.

This then completes the analysis of the validity results initiated in this chapter.

4.9 Conclusions

In summary, this chapter has basically dealt with the validity problem, with a specific reference to the criterion-related validity within the frame work of internal continuous evaluation system at the tertiary level. For this analysis, the examinations at BITS have been selected as the data base.

To pursue the criterion-related validity study, following assumptions were made:

1. The objective of the validity study is taken to be to see how well a student would do in his professional life.

2. Apart from examinations like Higher Secondary, performances in courses from Physics and Mathematics disciplines are considered to be central to the preparation of the student for the above objective. Thus, performances in courses like Modern Physics and Concepts in Science are taken to be good candidates for being predictors.

3. The choice of the various criteria is made from different assessments as reflected through GPA BIO, GPA CHEM, GPA MATH, GPA PHY, GPA SCI and CGPA.

4. All the campus based examinations are assumed to have similar structure and objectives.

5. The study has selected 1976 input of 310 students and 1977 input of 316 students as the data-base. For the various purposes of the analysis a temporal invariance is assumed in the general character of these samples.

As indicated above, this chapter exploits the data base as obtained from the 1976 and 1977 inputs. The observations emerging from the validity study as conducted in this chapter for the above representative student sample are as follows:

It is interesting to note that PHY GPA and CGPA are emerging as the best criteria against the Modern Physics as predictor for each of the disciplines or the groups of disciplines considered for the study. Further the above criteria-predictor pairs are more suitable for the Science and Applied Science & Technique oriented groups than for Engineering group, in that order. And further analysing these two pairs with respect to the Engineering group, Modern Physics - CGPA emerges as a slightly better pair than the Modern Physics - PHY GPA pair.

While studying the concurrent validity it has been found that for all the samples except in the case of the sample for ME students, the correlations between the performances pertaining to the GPA in ENGG courses (predictor) and GPA in their own disciplines (criterion) show correlation levels between moderate to high. Thus, in general, the performances within the institute based courses seems to be consistent, thus giving a good concurrent validity.

The above then is the total frame of reference under which validity analysis has been studied for examinations under internal continuous evaluation system as available through the field study at BITS.

Some suggestions for futuristic study emerge out of this investigation. To be specific, the cases where correlations between examinations are poor need to be explored further in a detailed and structured manner so as to study the test reliability, content validity as also the content correlation in order to understand why the correlations are poor.

This then summarises the observations of this chapter. Like the internal system of continuous evaluation and integrated approach to education, another major innovation at BITS is in terms of its efforts to link the education with the environment through its Practice School system of education. In the chapters to follow this thesis would investigate some aspects of the reliability and validity for the Practice School system of examination.

CHAPTER 5

ON THE PRACTICE SCHOOL SYSTEM OF EDUCATION - A CASE OF MULTIPLE OBJECTIVE EXAMINATIONS

5.1 Introduction

So far the thesis has discussed the reliability and validity aspects of the campus based class-room courses. As indicated earlier, this chapter initiates the study of the examinations administered under the Practice School (PS) System of education which essentially links the system of the university education with the real-life problem solving involvements of the professional world. Towards this, because of its (BITS) acceptance as early as 1970 of the PS system of education as an integral part of the student training for a degree across any branch of the Engineering, Sciences and Humanities disciplines, understandably, as in case of the previous chapters, this chapter also, extensively, draws upon the vast data base available in this context at BITS, Pilani.

The PS course is totally different from the conventional class-room type courses. It requires the involvement of students in real-life projects and is held not in a university campus but in professional surroundings like an industrial organisation, research laboratories, banks, etc. Some details of the PS education are given in section 5.2 of this chapter.

The PS education, in view of its close linkage with the professional world, makes certain new demands on the university population - both teachers and students - quite often in addition to those demanded by the class-room education. In this system both teachers and students become part of a team, being learners, educators, managers all at one and the same time. Students, on their part, have to accept certain consumer obligations whereas teachers have to appreciate and accept a new role. This theme has been elaborated in section 5.3.

As discussed in section 5.4, the usual techniques of examination will not be applicable to the PS course. In contrast to the class-room evaluation, the PS evaluation is multiple objective type. The PS evaluation method and its multiple objective nature is discussed in section 5.4.

5.2 Practice School Education

The PS education has been described in a series of articles by Mandke (1975, 1976, 1978) and BITS bulletin (1980-81), and also by Mitra (1974, 1975). A comprehensive monograph by Mandke (1980) gives the philosophy, history and operational details of PS alongwith a critical analysis of its impact on the university system. In this section have been included only a brief description of the PS system and its educational implications. For details the aforementioned references may be consulted.

As emphasized by Mandke in a series of articles mentioned above, PS serves an important need of the university education system. To appreciate this fact one can consider the aims of higher education as idenfified in the Carnegie Commission report (1973) quoted below:

1. The provision of opportunities for the intellectual, aesthetic, ethical and skilled development of individual students and the provision of campus environment which can constructively assist students in their more general development and growth.

2. The advancement of human capability in society at large.

3. The enlargement of educational justice for the post-secondary age group.

4. The transmission and advancement of learning and wisdom.

5. The critical evaluation of society through individual thought and persuasion for the sake of society's self-renewal.

In a nutshell, the aim of higher education is to make ~~a~~ better and more capable human beings who can serve the diverse social needs. However, it is obvious that due to a large variety of specific needs which keep changing from time to time it is not possible for the conventional class-room type education system to train a student for all such social needs. This is what is referred to by Jevons (1972) as "----- The clear impossibility of providing custom-built training for the infinite variety of job contents that exist at present, let alone those that will come to exist over the next thirty to forty years".

What is actually needed therefore is to train the students in the techniques of problem solving. It certainly requires knowledge which, as we saw in chapter 2, is one of the important objectives of education and which, to a fair extent, can be achieved in the class-room. However, the usefulness of knowledge imparted in the class-room can be quite limited as observed many years ago by Whitehead (1933):

"In my work at universities I have been much struck by the paralysis of thought induced in pupils by the aimless accumulation of precise knowledge, inert and unutilised. In a sense, knowledge shrinks as wisdom grows, for details are swallowed up in principles. The details of knowledge which are important will be picked up adhoc in each avocation of life, but the habit of the active utilisation of well understood principles is the final possession of wisdom."

The useless accumulation of inert knowledge does not make a student useful for society. Based upon this, Bayley (1972) suggests,"since knowledge is important as it is used, tasks would have to be set for the student requiring activity beyond sheer retention."

Bayley (1972) further writes, "Teachers have long bemoaned the absence of 'synthesis' or intellectual "integration in higher education, meaning that students do not see the interrelationships between bodies of knowledge. The solution has customarily been interdisciplinary courses and seminars in which subjects have been juxtaposed for the student. That such attempts have failed is not surprising. Synthesis occurs in human mind; it occurs when an individual discovers that academic delineations of subject matter break down in practice. Interdisciplinary courses fail because they are constructed externally to the student, interdisciplinary offerings become simply other courses produced for consumption"

The point that one wishes to make here is that although the class-room education can provide knowledge yet something else is needed to make this knowledge useful in solving real-life problems. Real-life problems are interdisciplinary in nature. As Bayley (1972) points out in the words quoted above, the solution would not lie in terms of merely offering interdisciplinary courses at the universities. The students have to be trained in the techniques of synthesising knowledge.

It is the lack of appreciation of these facts which has resulted in the universities imparting merely what

Whitehead calls 'inert knowledge' to the students. This has made the universities develop into directions which, as put by Mandke (1980), do not contribute to social growth. The irrelevance of university education to social needs has resulted in what is described by Mandke (1975) as 'the crisis of higher education'. In the words of Mandke, "----the (Indian) universities have not been able to integrate their activities with those of national life----higher education drinks from the well of society and hence they (universities) must turn their faces to the environment-----. The need of the hour is to evolve our own model which will help us in demystifying higher education by making it the summation of the people's experience so that it can properly belong to the society. One way to meet this requirement is to adopt the Practice School System of education".

The Practice School can be described as "a controlled simulation of the professional life during the educational years". It provides a platform for the interaction of the university with environment which eventually makes an impact on both the university and the environment. In real terms, a part of students' educational programme is carried out at a PS station which may be in an industry, a bank, a research laboratory or even a village. Students work on projects which are identified at the PS station and are naturally of direct interest to the 'host'. There are two PS courses PS-I and PS-II in the entire 5-year period with reference to the 11 year input. Consistent with the recommendations of AICTE, the duration is now changed to 4 year for the 10+2 input. However, duration and the locations for the PS courses in terms of time and structure, respectively, remain unchanged. PS-I is for one summer where as PS-II is for about six months which is slightly longer than one semester. Students work on projects in teams with one of them identified as leader for each team.

Students' performances in both the courses, PS-I and PS-II are evaluated in terms of letter grades like most other courses at the Institute. The grades are awarded by the Institute faculty who supervise the student work at the PS station. It is the physical presence of the faculty at the PS station and the fact that it is the Institute faculty and not the host organisation's staff that take the responsibility of organising and running the PS station and evaluating the students' performance which constitutes a vital feature of the Practice School. A measure of the success of the Practice School system of education is its acceptance by the students is evident from the fact that although it is not compulsory almost all the students of BITS invariably choose the Practice School option (see Mandke (1980) for full statistics). Thus, in any given academic year, out of the total of the student and faculty strength of around 2000 and 200, respectively, as many as 800 students and 60 faculty members are participating in the PS education at about 30 round-the-year PS-II stations and 45 summer-based PS-I stations situated at in all around 25 different towns/cities of the country. The real strength of the acceptance by the host organisations of this programme is indicated by the fact that almost all students particularly at PS-II station receive some financial contribution from the host organisations in terms of out-of-pocket allowance or subsidised food and accommodation (see Mandke, 1980 for details).

As is apparent from what has been said above, Practice school has made a major impact on BITS and its success has brought new obligations and responsibilities to both teachers and students of the Institute. This aspect is discussed in greater detail in the section that follows.

5.3 Changing Role of Faculty and Students - Consumer Obligations

The role of the faculty and students in a conventional education system is very well defined and has been accepted by tradition. A teacher has to teach and evaluate whereas a student has to learn what is taught to him and reproduce the same as faithfully as he can in the examination. In the Practice School system neither of these descriptions of the roles is valid as has been amply shown by Mandke(1980). This section takes into account the changing role of both the teachers and students in the Practice School system of education.

According to Furst (1958) "Every teacher and every college should formulate and use a defensible theory of learning". This will necessarily require search for a set of specific aims and objectives of education. According to Heywood (1977), "The derivation of objectives not only involves us (teachers) in an understanding of our philosophical position but in the formulation of a defensible theory of learning".

The importance of identification of the teacher's personal philosophy has also been emphasised by Sherin and Long (1972) who write : "If we want Clyde (name of an imaginary student) to learn what we must teach him, we must:

- be able to rationally communicate to him, our reasons for certain instructional goals and objectives;
- Consider Clyde's goals and objectives as we search for our optimum instructional system;
- be able to convince Clyde that it is important to learn certain behaviours, attitudes and skills which will become his "Professional Characteristics" as he interacts with society in the practice of his profession;
- be able to understand the value of the subject we offer for the fulfilment of Clyde's purposes, our purposes and the purposes of society".

A more rigorous and comprehensive discussion of aims and objectives in relation to the PS education and the corresponding 'defensible theory of learning' is found in Mandke's series of articles quoted in the beginning of the preceeding section. The first step for the teachers is to formulate, what Heywood (1977) and Sherrin and Long (1972) call, their 'personal philosophy' in the context of the PS education. This is important for the teachers in order to identify and play successfully their role in this novel educational innovation which, inspite of its initial success, is yet to be blessed by tradition.

The role of faculty in PS is indeed drastically different from that in the traditional class-room education where the teachers teach and evaluate students' performance in the subject of their specialisation. In PS, a faculty member does not teach his subject of specialisation but is along with the students a member of the team involved in the execution of a real-life project. As described by Mandke (1980), "---- the PS theory of higher education and research is based on the central theme of the teacher-student participation in the environment".

It is important to recall that the real life problems are usually interdisciplinary. The students in the team would also be from different branches. A teacher can not be an expert or specialist in all disciplines. He is therefore a member of an interdisciplinary team working on a problem of which the answer is not known either to the teacher or for that matter to anybody else. Indeed, there is no guarantee that the answer even exists.

And compare this with the class-room situation where the teacher is an expert in his subject which he expounds to his students. Here indeed, the correct answer to any problem is known to the teacher. In fact, Mandke (1980) himself is provoked to ask "---- what is the exact role of a teacher when students are working on a development type activity (like can be the case at PS) ? What teaching techniques should be adopted when discussing problems of which the teacher himself does not know answers ? Can a language teacher evaluate an engineering student working on, say, some feasibility study ?"

The answers to the above questions come atleast partly from the following. Although accepted by tradition, there is really no need for the teacher to serve the role of a living library packed with complete information about what he is expected to teach. If the job of the teacher were simply to transmit the existing information to the students, it could be achieved much more efficiently by some well written books, video tapes, etc. In such a situation, only a few teachers of excellence would suffice to cater to the whole world. Since the aim of education afterall is to produce personnel for solving real-life problems, the role of the teachers should be to train the student in the art and science of tackling and responding to real-life problems.

It should be emphasised that it does not undermine the importance of specialisation of teachers in particular disciplines. A Physics teacher must be a specialist of Physics but he should be able to communicate his problem-solving expertise and methodology to tackle a real-life problem which may involve knowledge of, say, biosciences or chemistry or chemical engineering or even of finance.

This also emphasises the importance of developing a suitable personal philosophy by the teachers. A bio-science teacher, who is the member of an interdisciplinary team in PS, should not feel embarrassed because his knowledge of thermodynamics is less than, that of a student-member of the team who for example, may be a student of chemical engineering. The teacher would accept and respect this difference and guide the student in how to use his knowledge of thermodynamics to respond to a particular problem.

In addition, the teacher has to inculcate the proper attitudes in his students. He has to combine the academic rigour with professional ethics and discipline and transfer this to the students and inspire them to face real-life challenges. He also has to plan suitably the deployment of the student manpower. He is not an expert in all subjects but he would know the expertise of others. In other words, the teacher has also to play the role of a manager.

To summarise : the changing role of teachers can best be explained by making the following quotation from Mandke (1980), "(these issues) relate to the new professional challenges that the PS system of education puts before the teachers------the teacher acts as the most important link between theory and practice----the faculty has the task of evolving teaching and evaluation techniques which are 'performance' based in the sense they facilitate, against the concept of the delivery schedule, the periodic assessment of the problem-solving tasks at hand as well as of the multi-objective contribution towards the same by the individual members of the student team. This 'project'-oriented approach to the teaching and learning processes is certainly a major departure from the usual university class-room norms".

Obviously, the methods of evaluation in this context are extremely important. A biosciences teacher in the above example could not and would not attempt to evaluate the expertise of students in for instance heat transfer or thermodynamics. What he would evaluate is the personality traits and skills required for solving real-life problems? What he would then try to evaluate is, say, how the students have responded to problem solving situations ? In view of its significance, the PS evaluation will be taken up again for further discussion in the next section.

The present thesis now considers the role and obligation of students in the PS system of education. In this system the students are not merely learners. They make an impact on the system by contributing to the host organisation in terms of assisting them as interns in their work and to the university by giving it an opportunity to bring the real life environment in to the class-room. The students, like the teachers, are an essential part of the real life problem solving network and thus have a role to play in this network. This also puts them under certain obligations which, in the words of Heywood (1977) may be called the "consumer obligations".

Heywood defines the "consumer obligation" for the student as follows:

"Just as the teacher has obligations in respect of students which require him to develop defensible theories of knowing and learning, so the student has obligations to acquire skills of learning and recognise impediments in their attainment".

Heywood (1977) further observes, "----it (higher education) is very much at the mercy of the consumer (and) a major outcome of such processes ~~(ie, conflicts between the obligations)~~ is that students may reject perfectly laudable activities if they do not seem to meet the needs of their future life as perceived by them at that time".

The experience of the PS so far clearly shows that it has been accepted by the student community. An indication of the students' attitude was reflected the fact that in order to go to PS they were even prepared to stay in road-side tents (Mandke, 1980) in the absence of any other arrangement for their stay Consistent with the statement of Heywood (1977) quoted above, this shows that this programme certainly meets the needs of their future life as perceived by them at the moment. According to Mandke (1980) the student community has accepted PS for another, perhaps more profound, reason which is that they 'experienced the excitement and satisfaction that comes when one transforms knowledge into performance, when one blends subjects learnt into techniques of action'. According to Mandke (1980), again the student plays the role of a learner and an understudy or some kind of a junior manager in PS. The acceptance of such a role has strongly contributed to the growth of PS. The students have indeed acted as 'catalysts and predictors of this educational innovation'.

The role which the students play in the PS is obviously very different from their role in the conventional class-room education where they simply have to be passive learners. Indeed, P.S. gives them an opportunity, in addition to learning a subject, to develop what Heywood refers to as the 'skills of learning'. A direct involvement in a real-life project brings out certain personality traits which have no chance to surface in the conventional class-room situation. These traits are the qualities such as leadership of a team, adherence to professional discipline, planning and sticking to a time schedule, etc.

Once a project is taken up by a team of students, it has to be tackled within a certain time schedule. Obviously, the solution would not be known to any-body beforehand including the teacher himself. The problems may not be fully categorised in terms of any one particular discipline and would, in general, call for a multidisciplinary approach. In an attempt to solve the problems associated with the project, the students may have to interact with the personnel of the host organisation and therefore learn and manage to establish appropriate human-relationships a factor of great educational importance. Thus, in the PS system a student has to play the role of not merely a passive learner but of a junior manager.

All this requires the so called 'consumer obligations' on the part of a student. First of all a student has to accept his aforementioned role. He must adhere to the discipline and rules and regulations of the host organisations even if formally he is not an employee of that organisation. Further, another important obligation on the part of students would be to mentally accept the fact that the role of his teacher in PS is different from that in campus-based class-room

The teacher in the PS, in general, would not be the 'expert' who has with him the solutions to all problems associated with the project. Indeed as observed earlier, the solution of these problems in certain cases may not even exist.

This completes observations on the changing role of students and faculty under the PS system of education. A natural philosophy to the above is the observation that understandably, then, the student evaluation at PS is multi-objective in character. Next section is devoted to this character of the PS examination system.

5.4 Practice School Evaluation and its Multiple Objective Nature

Since the teaching and learning in PS is quite different from the conventional class-room education, the techniques of evaluation at PS have also to be different. The class-room examinations, as has been seen in chapter 2, mostly test knowledge and ability to recall. The PS examination techniques are "performance based in the sense that they facilitate the periodic assessment of the problem-solving tasks at hand as well as of the multi-objective contribution towards the same by the individual members of the student team". (Mandke 1980).

By its very nature, the PS evaluation, unlike the class-room evaluation, cannot be based upon the single objective of testing knowledge. The fact that the PS takes the education of a student outside the class-room requires that the examination reform be also taken outside the class-room.

It is interesting to see the PS evaluation in the context of the system of internal continuous evaluation. According to Mandke (1980) the culture of the internal system of continuous evaluation matches well with the industrial practices of periodic assessment of the tasks at hand as well as the performance of manpower. It is therefore natural that the same be adopted to suit the evaluation of the students' performance at the PS.

Keeping in view the demands of a real-life problem-solving effort, the PS evaluation methods must test the students' performance in terms of personality traits like ~~[illegible]~~ professional judgement, decision-making ability, skills in data handling, ability for written and oral presentation, initiative, ability for team work, leadership quality, ability for meeting deadlines etc." (Mandke, 1980).

Thus we see that in contrast to the uniobjective nature of the class-room evaluation, the evaluation at PS is multi-objective. An evaluation system which attempts to achieve these objectives has been developed by the PS division at BITS. This system is based on the use of such components as seminars, quizzes, viva, group discussions, direct observation of the students, etc. This evaluation scheme is quoted here in a matrix form in appendix A glance at this will clearly show the students' characteristics which the scheme attempts to assess and the various instruments of evaluation used for assessing these characteristics.

An important point to note, as emphasised by Iyer and Krishnamurthy (1976), is that the PS evaluation is possible only in an internal continuous evaluation system, which also matches with the Industry's own system of assessment (Mandke, 1980). As such, the PS should be considered as a part of the Institute's broad-based interdisciplinary integrated educational system. The reliability and validity of class-room courses in an internal continuous evaluation system have been discussed in the previous chapter. It is only natural, then, that, against the above mentioned framework, the next chapter then, be devoted to the reliability and validity analysis for the PS examinations. As indicated earlier the validity analysis in the next chapter centres around the theme of the content validity.

CHAPTER 6

Some Aspects of Reliability and Validity Studies for the Practice School Examinations

6.1 Introduction

A brief description of the Practice School (PS) system of education, along with the statement of some of the salient features of its examination system, is presented in the previous chapter. This chapter, therefore, concerns itself with certain reliability and validity studies pertaining to the PS examinations.

More specifically, sections through 6.2 to 6.5 discuss certain reliability analysis techniques for PS evaluation. As one is aware, PS evaluation consists of components such as quiz, viva, seminar, group-discussion, project-report, observation, etc. Thus, section 6.2 deals with the problem of the study of reliability of PS evaluation components of quiz and viva, while section 6.3 is concerned with the problem of the study of the examiner reliability with reference to the PS evaluation components of seminar, group-discussion, project-report and observation. As regards to section 6.4, it pursues the theme of the course reliability with reference to the PS course, while section 6.5 deals with inter-correlations between the various above mentioned and any other components of the PS evaluation scheme. Against this, section 6.6 is devoted to the study of content validity of the PS evaluation. Finally, section 6.3 concludes the chapter.

It may be mentioned that while sections 6.2, 6.3 and 6.5 use marks obtained by students at Nagda PS-II station during the session I semester 1977-78 as the data-base for the purpose of the illustration of the emerging techniques, section 6.4 uses the data-base as available through Nagda PS-II sessions of I semester 1976-77 and I semester of 1977-78.

One reason for the choice of the above two sessions is obviously in terms of the availability of the adequate data-base for the purpose of different investigations to be pursued through the sections to follow. However, as regards to the choice of Nagda PS-II station in particular, it may be mentioned that it is because of the fact that, in terms of the educational infrastructural details, the matrix of the student-teacher interaction at Nagda station is intensively closer to the known class-room world within the campus-based university education.

6.2 On the Reliability of PS Evaluation Components of Viva and Quiz

A critical look at the PS evaluation scheme described in section 5.4 should tell that it (the PS evaluation scheme) is a peculiar mix of the traditional evaluation components adopted in the campus-based university system and the on-going personnel appraisal practices as operative in the industrial/professional set-ups. More specifically, the use of quizzes as a means to evaluate the student performance is an integral part of the evaluation system, particularly, in case of universities based on the semester pattern of education with continuous evaluation, while the practice of using viva, i.e., an oral interaction between the student/junior manager and the teacher/senior manager as a method to assess the progress/performance is common to both the academic as also the industrial/professional communities. Against this, the evaluation/assessment components like group-discussions, seminars, project reports, observation, day-by-day technical reporting of the work progress, etc. are uniquely closer to the personnel assessment patterns to be normally found in the modern and progressive industrial/professional environments of the twentieth century.

Coming to certain operational details pertaining to the PS evaluation scheme, it may be mentioned that, consistent with the requirement of the internal system of education, the PS evaluation certainly is the responsibility of the Institute's faculty manning the station. However, as the professional experts from the host-organisation along with the Institute faculty, play the role of the consultants to the students working as interns on the given PS-II assignment, the PS evaluation scheme, to the extent feasible, seeks the participation of the professional experts in the implementation of the various evaluation components.

For example, during the Nagda PS-II sessions, for which the field data is collected for some of the illustrations in this chapter, the professional experts heavily participated in the components of group-discussions and seminars. Thus, for these components, the experts ranked the students on the 0-10 scale and this information was used by the faculty as one facet of the total data-base before them to arrive at the final assessment of students for those components.

Similarly, the project reports were also, particularly, at the manuscript stages, invariably shown to the respective professional experts and their views were sought in the same context and these views did form one important base of information in terms of the final assessment of the project reports. Further, the evaluation for the 'observation' component also made use of the periodic feedback from the professional experts on the student progress and professional behaviour. Finally, in case of the component on the (technical) diary also the faculty sought a feedback from the experts as and when needed. The only PS evaluation components, where the involvement of the professional experts was to its minimum, were, thus, quiz and viva.

It is against the above frame of reference that firstly, in view of their strong identification with the student evaluation practices traditionally followed in the university system and, secondly, in view of their being conducted in the industrial/professional environment almost entirely by the university faculty alone, this section studies the theme of the reliability of evaluation components of quiz and viva when implemented as under the guidelines of the PS evaluation scheme.

As indicated earlier, the data-base selected for the above indicated study consists of the marks obtained by the group of 30 students at Nagda PS-II session during the I semester of 1977-78. This student group consisted of the final year students from different disciplines of chemical engineering.

civil engineering, Electrical & Electronics Engineering, Mechanical Engineering, and Management. The team of instructors who supervised their PS-II education had four teachers drawn from different disciplines of chemical engineering, electrical and electronics engineering and management and one of them was the instructor-in-charge of the team.

During the Nagda PS-II session under consideration in all 4 quizzes and 4 viva were conducted. Thus, to study the reliability of the quiz component, each of four quizzes can be considered as separate item and then, under the analysis of variance approach, the reliability of the quiz component can be obtained through the equation 3.18.

Similar method as above can also be adopted to obtain the reliability of the viva component for the data-base under consideration. The results so obtained for the components of the quiz and viva are reported in table 6.1.

Table 6.1

Reliability of Practice School Evaluation Components of viva and quiz

Examination	Number of Examinations conducted	Number of students	Coefficient of Reliability
Viva	4	30	0.73
Quiz	4	30	0.43

Indeed, the sample used for the results reported in the above table is too small for any generalization. But, if one was to still conclude then it seems to emerge that for the batch studied the reliability of the quiz component, which is unique to traditional student assessment techniques followed in the university system, was lower than that for the viva component, which, while being present in the university evaluation system, also has an acceptability in the professional world. However, for any well founded generalization, it would be necessary, in a futuristic context, to pursue this investigation, along the methodology presented above, for a larger number of PS-II courses conducted at different stations as also for a larger sample size in terms of student number and discipline.

6.3 On the Examiner Reliability For the Components of Practice School Evaluation

One of the major criticisms of examinations relates to variation in marking. Marks awarded by one examiner to the same script may differ from the marks awarded by the other examiner. This variation in marking may bring in error in measurement. The consistency in the marks awarded by different examiners in the same examination is measured in terms of the examiner reliability. Gullicksen (1950) defines it as the correlation between the marks of two examiners and observes that it gives the measure of examiner reliability.

A large number of studies in India as well as abroad have been done on this subject (see section 3.5). However, generally, such studies have been done on the conventional class-room tests. Thus, the data base available from the Nagda PS-II station in terms of the marks of the I semester 1977-78 batch, where each of the components of PS evaluation, namely, seminar, group-discussion, project report and observation has been evaluated at a time by a number of examiners their numbers varying between 2-4, provides an interesting case for an investigation with reference to the problem of the examiner reliability in an environment outside the traditional class-room culture.

As regards to the method, it is submitted that for any of the PS evaluation components under consideration, student evaluation by a given examiner can be treated as one separate item. That is to say, a seminar component examined individually by three different examiners can be treated as an examination or test with 3 items. In such case, the model relating the observed score with the true score can be same as the one represented through equation 3.18, with α_i representing the error in evaluation due to the factors concerning the examinee and the β_j representing the error in evaluation due to the inconsistency in marking by the examiners. Indeed, all other assumptions concerning the model as represented through equation 3.18 can also be taken to be valid in case of the details pertaining to the present investigation. Thus, the examiner reliability can, then, be calculated through the analysis of variance approach using equation 3.18.

It is against the above frame of reference that the table 6.2 presents the results on the examiner reliability for the PS evaluation components of seminar, group-discussion, project report and observation, when implemented for the I semester 1977-78 batch of Nagda PS-II students.

Table 6.2

Examiner Reliability for Different Practice School Evaluation Components as Implemented for I Semester 1977-78 batch of Nagda PS-II students

Examination	Number of students	Number of Examiners	Coefficient of Reliability
Seminar II	30	3	0.79
Seminar III	30	3	0.50
Group Discussion II	30	3	0.84
Group Discussion III	30	3	0.70
Project Report	30	4	0.68
Observation	30	2	0

From the results above reported, it clearly emerges that the PS evaluation components of seminar, group-discussion and project report, studied in terms of the data base as available for the I semester 1977-78 batch of Nagda PS-II students, demonstrate a fairly satisfactory level for the examiner reliability. More specifically, as can be seen from table 6.2, the examiner reliability in case of the two group-discussions studied is invariably on the higher side, the reliability coefficients being of the order of 0.84 and 0.70. Against this, the examiner reliability for the two seminar components is found to be of the order of 0.79 and 0.50. Thus, even though, on an average, the group-discussion component seems to have demonstrated a better examiner reliability than that for the seminar component, the examiner reliability for the later, i.e. the seminar component, is also fairly satisfactory, the lower value for it being as high as 0.50.

Finally, the project report component has also demonstrated a fairly satisfactory examiner reliability, its value having emerged as high as 0.68, which is only slightly higher than the average value of 0.645 for the examiner reliability for the seminar component. The reasons for such a satisfactory level of the examiner reliability in case of the PS evaluation components under discussion could be seen in terms of the reduction in the error (due to the examiner subjectivity) that seems to be feasible particularly in view of (i) ~~Firstly~~, the structured nature of these examination components as reflected through the PS evaluation matrix of Appendix C and, (ii) ~~secondly~~, the participation, in an involved manner, of the professional experts in the PS evaluation process.

However, in terms of the data-base as available for the I semester, 1977-78 batch of the Nagda PS-II students, the observation component of evaluation demonstrates 'zero' examiner reliability, thereby indicating in-consistency in the evaluation.

It is true that, as in the cases of the seminar, group-discussion and project report components, the above studied observation component also has a structured format as reflected through the PS Evaluation Matrix of Appendix C and, as described in section 6.2, the faculty, i.e. the examiners, while arriving at their student evaluation in terms of the observation component, also had adopted the practice of seeking the view of the professional experts. And, still, the examiner reliability for the observation component studied has come out to be zero. This, then, only goes to show that when it comes to the 'observation' component, in spite of all the checks and balances provided, finally, the assessment seems to be heavily weigh in terms of the perspective through which a given examiner is viewing the student; thus resulting in a possibility of a higher order of inconsistency in the evaluation. One way to reduce this inconsistency can be in terms of the examiner's training in the context of the various possible range of interpretations of human behavioural situations, so as to ensure a greater from-within debate in the mind of an examiner before arriving at his assessment of the student.

Thus, in summary, from the analysis of the results of the data base available for the I semester, 1977-78 for the Nagda PS-II station, it emerges that, in respect to the batch under consideration, the PS evaluation components of seminar, group-discussions and project report demonstrate a fairly satisfactory level of the examiner reliability. Further, the average examiner reliability for the group-discussion component is higher than that for the seminar component. As regards to the project report, its examiner reliability is seen to be lesser than that for both the group-discussion components studied and only slightly higher than the average examiner reliability value for the seminar component. Finally, the observation component is noted to have a very poor examiner reliability.

Indeed, the above sample size is too small for any generalisations. Towards this, further investigatory work along the methodology developed in this section, would be necessary in terms of a larger sample size which would need to be collected in a planned manner.

6.4 Course Reliability for a Practice School Course

Another interesting investigation that can be pursued with regards to the PS evaluation scheme concerns the question of consistency between its various components of evaluation, namely, quiz, viva, group-discussion, seminar, project report, observation and diary. This indeed is nothing but the study of the PS course reliability. Towards this, along the similar lines as in the case of course reliability for a campus-based course described in section 3.4, a given evaluation scheme for a PS course can be considered to be a semester long examination, with each of the above listed seven evaluation components signifying a separate item. ~~Thus, then, the observed score of a student can be described through a model:~~

The observed score of a student can then be described through a model represented by following equations:

$$x_{ij} = t_{ij} + e_{ij} \tag{6.1}$$

and

$$t_{ij} = \mu + \alpha_i + \beta_j \tag{6.2}$$

under the assumptions that:

(i) error components, e_{ij}s, occur independently and at random with zero mean and common variance, and

(ii) that they (i.e. e_{ij}s) are uncorrelated with t_{ij}s and with errors in other measurements,

where,

(a) $x_{ij} \triangleq$ Marks obtained by the i^{th} student in the j^{th} evaluation component.

(b) $\mu \triangleq$ Factor common to all measures (i.e. scores),

(c) $\alpha_i \stackrel{\Delta}{=}$ Factor in the total measure signifying the contribution of the i^{th} student,

(d) $\beta_j \stackrel{\Delta}{=}$ Factor in the total measure signifying the contribution of the j^{th} evaluation component.

and (e) $e_{ij} \stackrel{\Delta}{=}$ Unexplained factor or the error in the measure for the i^{th} student due to the j^{th} evaluation component.

Consistent with the treatment stated in section 3.4, the PS course reliability can, then, be obtained through the analysis of variance approach using equation 3.18.

The PS course reliability results emerging through the application of the above mentioned to the data-base available in terms of the marks obtained by the two batches of PS-II students at the Nagda PS-II station during the sessions I semester 1976-77 and the I semester, 1977-78 are reported below in table 6.3. The I semester 1977-78 batch has been in detail described earlier in section 6.2. As regards to the I semester 1976-77 batch, it may be mentioned that it consisted of single faculty which belonged to the mechanical engineering discipline and seven students pertaining to chemical, electrical and electronics engineering and management disciplines.

Table 6.3

Coefficient of Course Reliability for the Practice School-II Evaluation

S.No.	Semester & Year	Total student No. as also their discipline-wise composition	Total No. of items	PS-II Course Reliability
1.	I Semester 1976-77	8	7	0.75
2.	I Semester 1977-78	30	7	0.73

As can be seen from the above table, the coefficient of reliability for PS evaluation scheme seems to be quite high. One of the reasons for getting high reliability may be the highly structured nature of the PS evaluation scheme as also the existence of a large number of evaluation components.

6.5 Inter-Correlations between the Practice School Evaluation Components

As indicated in section 6.1, this section studies the inter-correlations between the various PS evaluation components of quiz, viva, group-discussion, project report, seminar, observation and diary. The methodology for obtaining the correlation coefficients between the different PS evaluation-component-pairs is same as the one described in section 3.2.2 of Chapter 3.

Tables 6.4 and 6.5 give the results on the correlation coefficients as obtained from the marks of the two batches of the PS-II students at the Nagda PS-II station during the sessions I semester, 1976-77 and I semester, 1977-78, respectively.

Table 6.4

Intercorrelation Between Different Components of PS-II Evaluation at Nagda - I Semester, 1976-77

Components	Quiz	Viva	Seminar	Group Discussion	Project report	Observation	Diary
Quiz		0.43	0.68	0.70	0.45	0	0
Viva			0.85	0.86	0.26	0.60	0
Seminar				0.92	0.26	0.29	0
Group Discussion					0.53	0.57	0
Project Report						0.61	0
Observation							0
Diary							

Table 6.5

Intercorrelation between Different Components of PS-II Evaluation at Nagda - I Semester, 1977-78

Components	Quiz	Viva	Seminar	Group Discussion	Project Report	Observation	Diary
Quiz		0.24	0.50	0.32	0.57	0.06	0.14
Viva			0.18	0.22	0.19	0.14	0.10
Seminar				0.78	0.36	0.32	0.25
Group Discussion					0.26	0.54	0.27
Project Report						0.15	0.38
Observation							0
Diary							

The analysis of the results reported in above tables is presented below in table 6.6

TABLE 6.6

Evaluation Component-Category-wise Analysis of the Numbers of the Evaluation-Component-Pairs as coming Under Various Correlation Levels

S. No.	PS Evaluation Component Pair	No. of PS Evaluation-Component-Pairs as coming under a given correlation level					
		Zero	+ve slight	+ve Low	+ve Moderate	+ve High	+ve Very High
1.	Q-V			1	1		
2.	Q-S				2		
3.	Q-GD			1	1		
4.	Q-PR				2		
5.	Q-Ob	1	1				
6.	Q-D	1	1				
7.	V-S		1			1	
8.	V-GD			1		1	
9.	V-PR		1	1			
10.	V-Ob		1		1		
11.	V-D	1	1				
12.	S-GD					1	1
13.	S-PR			2			
14.	S-Ob			2			
15.	S-D	1		1			
16.	GD-PR			1	1		
17.	GD-Ob				2		
18.	GD-D	1		1			
19.	PR-Ob		1		1		
20.	PR-D	1		1			
21.	Ob-D	2					
	TOTAL	8	7	12	11	3	1

Below are listed some of the observations emerging from the table 6.6.

(i) There is no instance of negative correlation.

(ii) Little over 50% of the correlations between various pairs of PS evaluation components lie within the levels moderate to low.

(iii) About 35.7% of the correlations are within levels very high to moderate, while the remaining 64.3% are less than moderate.

(iv) The percentage-wise distribution of the correlation coefficients as coming under the various levels of Very High, High, Moderate, Low, Slight and Zero correlation is as follows:

Table 6.7

Percentage-wise Distribution of the Correlation Coefficients Coming Under Various Levels of Coefficient Correlation

Correlation Level	Zero Correlation	Slight	Low	Moderate	High	Very High
Percentage of PS Evaluation Component Pairs	19.05%	16.67%	28.57%	26.19%	7.14%	2.38%

(v) The seminar-group discussion pair has demonstrated the best correlations, while, on the whole, the diary component of PS evaluation shows indeed very poor correlation with all the other evaluation components.

(vi) On the whole, the pairs Q-S, Q-PR and GD-Ob show moderate correlations, while the pairs S-PR and S-Ob have low correlations.

The above then are some of the observations that emerge from the analysis of the inter-correlations between the various PS evaluation components as studied for the Nagda PS-II sessions of I semester, 1976-77 and I semester, 1977-78. Indeed, this sample size is small for any long term generalisations in terms of the above observations. Towards this, it would be necessary, in a futuristic context, to study along the methodology indicated above a larger sample on PS evaluation collected in a planned manner.

6.6 On the Content Validity of the Practice School Evaluation

The significance or the need for the validity analysis for an examination, along with the description of the types of validity, has been in detail discussed through section 4.2 of Chapter 4. Further, Chapter 4 has also investigated the predictive validity and the concurrent validity in terms of a series of traditionally administered examinations drawn from different levels of education as reflected through the Higher Secondary stage as also through the stages corresponding to the different levels of the academic years at the Institute. This section, thus, pursues the theme of the 'content' validity with reference to the PS evaluation.

6.6.1 Content Validation Methodology

Below is briefly stated the methodology for establishing the content validity of an examination.

One is aware that an examination is always conducted against a syllabus or a topic, as the case may be. Thus, the first step in the content validation of an examination is in terms of defining a specification matrix describing various educational objectives aimed at through the different topics or the sub-topics of the syllabus against which the examination has been conducted. Understandably, this task of identifying the content vs. educational-objective-based specification matrix is invariably pursued through the experts for the subject area from which the examination syllabus is drawn.

The next step, then, is to work out with the help of the examining faculty another specification matrix describing for each of the implemented items of the examination, the educational objectives for which the item has been designed. This matrix may be termed as the item-vs-educational-objective-based specification matrix.

The content validity for the examination under consideration is then defined as the correlation between the two above described specification matrices. In specific terms, higher the matching between the educational objectives as reflected through the two matrices, higher the content validity of the examination implemented (Anastasi, 1976).

From the above it follows as a natural corellary that, if at the time of the setting the question paper itself the examiner has the content-vs-educational-objective-based specification matrix before him and, as a result, if he so designs the test items or questions in the examination as to cover the entire range of the educational objectives stated in the above indicated matrix, then the content validity of the examination is automatically ensured.

The above corollary describes a situation where an examiner is setting a question paper for a given topic. If one considers a case, where students are being evaluated for a course, an amended version of the above corollary emerges.

For the purpose of the specificity, let the student evaluation in the course be on a continuous basis, thus implying presence of a series of tests or examinations at a periodic interval, ultimately resulting in a final student assessment. As can be appreciated, these tests or examinations can be in terms of various evaluation components like quiz, viva, tests, comprehensive, project report, home assignment, seminar, etc. Further, at the beginning of the course, let the evaluation-component-vs-educational-objective-based specification matrix be designed, matching with the content-vs-educational-objective-based specification matrix for the course.

The content validity for the evaluation scheme for the course is then automatically ensured if each of the examinations for the course is so designed and administered as to ensure its complete faithfulness to the details of the above stated evaluation-component-vs-educational-objective-based specification matrix.

It is against the above frame of reference that the next section discusses the content validation of the PS evaluation scheme as described through section 5.4 of Chapter 5.

6.6.2 Content Validation Analysis for the Practice School Evaluation

One way to study the content validity of the PS evaluation along the methodology described above could be in terms of a critical analysis of the historical background of the evaluation of the PS evaluation scheme. An excellent comprehensive account of this historical background is given by Mandke (1980) in his Case Study on the PS System of Education as evolved and implemented by BITS, Pilani. The contents of this section heavily draw from this case study.

As brought out by Mandke (1980), the PS examination system at BITS has evolved as a process, the starting point for which, indeed, is the Forward Plan of 1970 which for the first time defined in clear terms the educational objectives of the Co-operative Education as proposed to be implemented by BITS. Thus during the period 1970-72, then, there had been innumerable rounds of discussions between the PS Director from MIT (USA) and his MIT colleagues (i.e., the experts in PS education), the PS Coordinator at BITS, the proposed PS faculty for the first PS-I station at HCP (Khetri) and the first PS-II station at HINDALCO(Renukoot), the professional experts from HINDALCO and HCP and other senior officers of the Institute. In fact, during the summer of 1971, six faculty members, who were later to form the teams of faculty to man the first PS-I and PS-II stations, spent the summer of 1971 at HINDALCO to workout the entire range of the educational and operational details of the PS programme.

It was as a result of all the above mentioned deliberations involving people from University world as also from profession, along with the experts in PS education, that the basis got formulated for relating the contents of PS education as reflected through the structure and types of PS assignments with the objectives of the PS system of education. This basis, in the words of Mandke (1980), can be stated as follows:

"As a process of education, the practice method substitute the narrative approach followed within the four walls of a classroom and the two covers of the textbook by experience-based cognitive process of learning and teaching, operative in the very way of life, thus making education student centred and environment as well as circumstance oriented. In view of this the attempt in the Practice School is not merely to further student's knowledge in given codified orthodoxy (as this facet of education is well achieved in the classrooms), but to train him in the art of effectively contributing to real-life problem-solving efforts of production, of design, and ultimately of social action. Therefore, basically practice method aims at cultivating amongst students appropriate attitudes and analytical skills towards decision making and team spirit, responsibility of leadership, importance of time schedule and regularity, skills of written and oral presentations, organisational ability, etc. so necessary to respond to open-ended professional situations, more often than not characterized by insufficient data, uncertainties of events and unfamiliarity with the environment. Seen from this angle, it follows that though Practice School pursues project method of education, the attempt here is not to rediscover the past, but to use the contemporary day to day developmental activities of direct interest to the professional world as a vehicle to prepare students to participate in interdisciplinary, goal and mission oriented, and time-bound problem solving tasks."

Viewed in a broader perspective, the above extract, in a sense, can then be considered to constitute a kind of essay type analogue of the theme of the content-vs-educational-objective-based specification matrix as developed in section 6.6.1.

It may be mentioned that, the above deliberations also resulted in evolving the broad guidelines for the PS system of examinations. These guidelines are stated in Clause 22.2 of the academic regulations (1977) of the Institute quoted in Appendix H.

Indeed, it is in terms of the educational objectives of the PS system of education as reflected through the clause 22.2 of the academic regulations, that the first typical PS-I evaluation scheme was evolved and implemented in the summer of 1973 along with the first typical PS-II evaluation scheme that was evolved and implemented over years from 1973-1975 (See Appendix F.) It may be noted that at this stage of the evolution the PS evaluation scheme (particularly with reference to PS-I) ~~the scheme~~ had different formats for the engineering, science and humanities streams. Descriptions of some of these typical formats are given in Appendices D to E. It may be mentioned that based on the feedback available from the summer of 1973, the PS-I evaluation scheme was further modified in the summer of 1974 ~~and the same is given in Appendix~~ . At this stage it is important to recall that the faculty teams who had evolved and implemented these PS evaluation schemes were the same as those who had participated in formulating the basis for PS educational contents as also its examination objectives as reflected through the clause 22.2 of the Academic Regulations of 1977.

Next, around August, 1975, an extensive feedback was organised from students in terms of PS-evaluation. This feedback was sought through personal discussions as also through the intensely deliberated meetings. During one such crucial meeting, to begin with there was a suggestion to give more weightage to low level skills from the cognitive domain comprising of educational objectives like knowledge, memory, etc. However, as the deliberations proceeded, there turned out to be an unanimous support and acceptance of the point of view that the PS evaluation give greater weightage to the higher level skills from the cognitive domain like the ability to evaluate, decision-making, etc. Further, there also was a definite acceptance of the theme that the PS evaluation should assess the student professional personality traits in terms of qualities such as sense of responsibility, industry, self-expression, etc; thus further confirming the purport of the basis for the PS evaluation scheme evolved well before the starting of the PS courses.

As a natural consequence of the above, a study team was appointed to evolve a uniform PS evaluation matrix across the different disciplines. All the members of this team were equipped with past experience in PS evaluation. Applying its mind particularly to the task of designing the PS-II evaluation scheme, based on their experience as also based on the student feedback, this study team, within the frame of reference of the PS evaluation guidelines, prepared its first working paper in October, 1975, describing a PS-II evaluation scheme alongwith a rating sheet format for the same. The working paper was then distributed across the Institute inviting suggestions. Taking this as a basis, the study team prepared a second draft which in November, 1975 was discussed in a meeting in which as many as 75 faculty members participated. As a result of this a third draft emerged which in a marathon meeting held in December, 1975 was further discussed to arrive at a final evaluation scheme as stated in Appendices C and G . In terms of the details of the section 6.6.1, this PS evaluation scheme given in Appendix C indeed, constitutes the evaluation-component-vs-educational-objective-based specification matrix for the PS-II course. Around June, 1976, Institute's PS Division constituted a PS Instruction Cell with, amongst other things, the responsibility to plan and monitor the implementation of PS evaluation components. As a part of this responsibility of its, the PS Instruction Cell then particularly with reference to the PS-II course, developed its own progress and monitoring reports, periodically seeking from the PS-II faculty the information on the conducting of the PS-II evaluation components, so as to ensure their (evaluation components') matching with the contents of the evaluation-component-vs-educational-objective-based specification matrix of the Appendix C . In this context, the PS Instruction Cell also prepared a detailed implementational guidelines for the PS-II evaluation scheme.

Further, simultaneously, the PS Instruction Cell also constituted its own study team to vigorously interact with the PS-I students and faculty, so as to arrive at a uniform evaluation-component-vs-educational-objective-based specification matrix for the PS-I course consistent with the educational objectives pursued by the PS system of education. It was in the academic year 1977-78 that this matrix emerged in its final form. ~~and the same is reported in Appendix E~~ . Further, as in the case of the PS-II evaluation, the PS Instruction Cell also prepared a detailed implementational guidelines for the PS-I evaluation scheme. The above matrix in a checklist form and the above guidelines are presented in the PS Division publication titled "Guidelines for PS-I Operation and Evaluation" (1979).

It may be noted at this stage that, side by side to the task of finalizing the evaluation-component-vs-educational-objective-based specification matrix, the PS Instruction Cell also initiated efforts in the direction of consolidating the experience gained over years, so as to evolve a PS-I progress and monitoring mechanism (as reflected through the PS-I register) with a view to ensure the matching of the PS-I evaluation components implemented with the contents of the specification matrix of Appendix E .

Finally, by the year 1977-78, the PS Instruction Cell also regularised the process of seeking feedback from PS students in terms of administering of the survey type questionnaires as also through personal interaction, so as to further strengthen the monitoring process. Table 6.8 gives the analysis for the PS-II feedback questionnaires administered during the two PS-II sessions of the academic year 1977-78.

Thus, as can be seen from table 6.8, student response for both the batches to the query as to 'what was students' overall impression about the (PS) evaluation done by Practice Faculty' was overwhelmingly satisfactory. Specifically, if, from the I semester 1977-78 batch, as many as 76.3% of the PS-II students demonstrated satisfaction with the overall PS-II evaluation, for the II semester 1977-78 batch, the

TABLE 6.8

Results on Feedback on Evaluation from Students who attended Practice School in I and II Semester of the Academic Year, 1977-78

S.No.	Question	Semester/ Academic Year	Percentage of Responses			
			Yes	No	Not Responded	
1.	Is the evaluation Scheme too rigid?	I Sem.1977-78	28.9	53.9	17.2	
		II Sem.1977-78	47.2	50.0	2.8	
	What is your impression about the evaluation done by the PS Faculty?		Very much satisfied	Satisfied	Not Satisfied	Not Responded
2.	Quizzes/Tests	I Sem.1977-78	34.2	46.1	14.5	5.2
		II Sem.1977-78	41.7	30.6	5.6	22.1
3.	Seminars	I Sem.1977-78	28.9	51.3	15.8	4.0
		II Sem.1977-78	33.3	38.9	5.6	22.2
4.	Group Discussions	I Sem.1977-78	27.6	46.1	21.2	5.1
		II Sem.1977-78	25.0	33.3	13.9	27.8
5.	Viva/Oral Examination	I Sem.1977-78	27.6	47.4	14.5	10.5
		II Sem.1977-78	30.6	33.3	8.3	27.8
6.	Report Writing	I Sem.1977-78	32.9	46.1	15.8	5.2
		II Sem.1977-78	30.6	30.6	11.1	27.7
7.	Diary	I Sem.1977-78	27.6	42.1	23.7	6.6
		II Sem.1977-78	19.5	41.7	11.1	27.7
8.	Observation	I Sem.1977-78	32.9	35.5	26.3	5.3
		II Sem.1977-78	27.8	30.6	16.7	24.9
9.	What is your Overall impression about the evaluation done by Practice Faculty?	I Sem.1977-78	27.6	48.7	19.7	4.0
		II Sem.1977-78	30.6	33.3	16.7	19.4

corresponding percentage, though smaller, was also as high as 63.9%. Further, when enquired about their impressions relating to each of the PS-II evaluation components from quizzes to observation, then also majority of students from both the batches expressed satisfaction. In specific terms, if the above percentages for the I semester 1977-78 batch varied between 68.4% - 80.3%, the corresponding percentages for the II semester 1977-78 batch were between 58.3% - 72.3%. Finally, over 50% of the students from both the batches responded negatively when enquired if the PS-II evaluation scheme was too rigid.

Above then is a comprehensive description of the manner in which the design as also the implementational mechanism of the PS evaluation scheme as conducted at BITS have evolved. The main observations that emerge when the above details are abstracted are summarised below:

1. Practice School education system has evolved as a process. To start with, overall objectives were clearly stated; teachers from the university, experts from professional world and PS education experts participated in the whole planning process and finally arrived at a content-vs-educational-objective based specification matrix.

2. This was followed by finalisation of evaluation component-vs-objective based specification matrix by the same group. It may be pointed out that some of the members of the expert group themselves actively participated in the implementation of the PS evaluation scheme.

3. There exists a formalised Instruction Cell within the PS division with well structured methods to continuously get feedback, monitor and control the implementation of the PS evaluation scheme at different PS stations.

4. Instruction Cell also collects feedback about the PS Programme from the students on a regular basis. Sample survey of the feedback clearly showed that a good majority of the students are convinced about the credibility of PS evaluation scheme.

5. The very fact that professional experts were involved, while finalising the evaluation scheme as well as implementation of the scheme, not only ensured that all the important professional personality traits have been duly incorporated in the evaluation scheme but also ensured these are evaluated in a professional manner.

   Indeed the whole process at the same time ensured predictive validity as well as is evidenced by the great demand for PS transcript by the employers as well as students. The PS transcript gives details about the project assignments undertaken, grades obtained as well as a Rating Sheet which gives the faculty's rating on certain professional personality traits based on PS evaluation - which is presented in Appendix 4.6.

   Summing up, the very process of planning, scheduling, implementing and monitoring of PS evaluation scheme, within the framework of evaluation-component-vs-educational-objective based specification matrix, was such that it not only had all the prerequisites to ensure the content validity of PS evaluation scheme, but it also helped establishing credibility amongst students, acquired respectability among professionals, and, finally, achieved acceptance for itself from the academic community. It is against the above frame of reference that one can then see the content validation of the PS evaluation scheme described through this section.

6.7 Conclusions

This chapter has dealt with some aspects of the reliability and validity of PS evaluation by studying the data-base for

I semester 1976-77 and I semester 1977-78 as available from the Nagda PS-II station. The main conclusions of the study are as follows:

1. Practice School evaluation is a peculiar combination of the traditional evaluation components adapted in the campus-based university system as well as on-going personnel appraisal practices in the industrial/professional set-ups.

   Reliabilities for Quiz and Viva components of PS evaluation have been studied. It may be noted that Quiz as an evaluation component is predominant in the universities, Viva as evaluation component is common to both university system and professional world.

   The reliability of Quiz component was lower than the Viva component of PS evaluation - may be because of professional rigour that is brought in Viva component.

2. One of the major criticisms against traditional examination system relates to variation in marking by different examiners - an index of which is given by Inter-Examiner Reliability Coefficient. The same problem also exists in the professional world outside the class-room situation. It is in this background, examiner reliability is analysed for some of the PS-II evaluation components viz., Seminar, Group-discussions, Project report and Observation.

   For seminar, group discussion and Project Report the results are fairly satisfactory. One reason could be the possible reduction in error (due to examiner subjectivity), due to well laid out structure of the components. The other reason could be the effective participation of professional experts in the PS evaluation process.

However, the component 'Observation' showed 'zero' examiner reliability thereby indicating possible inconsistency in the evaluation. Thus inspite of highly structured evaluation scheme Observation demonstrated poor examiner reliability. This may be because the assessment seems to heavily weigh in terms of the perspective through which a given examiner is viewing the student, thus resulting in a possibility of a higher degree of inconsistency in the evaluation.

3. The coefficient of course reliability for PS evaluation scheme is found to be quite high. The main reasons for getting such high reliability are: (1) PS evaluation scheme is very well structured based on Evaluation component-vs-objective specification matrix, (2) the number of evaluation components are sufficiently large, (3) Effective participation of the professional experts in the implementation of the evaluation scheme.

4. All inter-correlations between various examinations of PS evaluation are positive. About 50% of the correlations lie within the levels low to moderate. A fairly good percentage (35.7%) of the correlations are within the level moderate to very high. More specifically, the seminar group-discussion pair demonstrated the best correlations while, on the whole, the diary showed very poor correlation with all the other components of evaluation.

5. To ensure content validity of PS evaluation scheme, a detailed specification matrix has been evolved defining various sub-aspects of each of the wide ranging educational objectives along with weights. Understandably, this has made the implementation of the evaluation system a physical reality.

It is against the above frame work, that one can confidently assert that the content validity of the PS evaluation is quite high.

6. It may be mentioned here that the sample selected for the study was small and hence it would be difficult to generalise the findings. However, the study clearly has shown the directions in which such research studies can be undertaken as well as focussed attention on some of the important aspects of PS evaluation. It would be worthwhile to pursue this thought process in a futuristic manner based on a larger data.

Above then are the broad conclusions as they emerge from the reliability as also validity based studies vis-a-vis the PS evaluation. As has been already indicated, the PS evaluation is essentially multiple objective in character, aiming at student assessment in terms of higher order educational objectives. It is in this context that one can then pursue the theme: whether the multiple objective evaluation that has been feasible under PS can be brought into the traditional class room based learning effort. The next chapter devotes itself to the purusal of this theme.

CHAPTER - 7

On Multiple Objective Examinations in Class-room Based Teaching at the Tertiary Level

7.1 Introduction

In this thesis the examinations have been discussed for the class-room based courses as also for the Practice School(PS) course. In the process, it is observed that the examinations in the traditional context as implemented through the class-room based system of university education are pre-dominantly measuring lower order objectives, while, under PS, the examinations are characterized by the multiple objective assessment of the student performance. As discussed in Chapter 5, this is possible in case of the practice method of education mainly because of the fact that in PS the teaching and learning as also the student assessment tasks are wholly conducted in the professional world.

The investigation to which this chapter devotes itself pertains to the query as to wether the traditional class-room based examinations be made multiple objective in nature. Towards this, section 7.2 of this chapter develops the theme of "transfer of learning", while section 7.3 goes to explore a model building exercise in the above context based on the experience gained through the educational innovation of M.E.(Collaborative) programmes recently evolved by BITS in direct collaboration with the professional world.

Finally, section 7.3 attempts to visualize the issues that should be emerging if the educational structure, as abstracted through the theme of the M.E.(Collaborative) programmes, were to be brought into the campus-based teaching activity as conducted in the traditional university set-up.

7.2 Transfer of Learning

An important reference to multiple-objective examinations in traditional class-room courses is due to Heywood (1977), who has studied this question in terms of the concept of "transfer of learning". It is in this context that the present section develops the theme of "transfer of learning".

Within the frame of reference of the demands of the contemporary world, the need for education to aim at training students to face the unknown can never be over emphasized. As rightly observed by Whitehead (1933) the theme of what can be called as "inert knowledge" has long served its purpose. It's place has now been taken by the requirement that students increasingly learn the skills for using 'knowledge' to solve real life problems. This has been referred to as 'transfer of learning' by Saupe (1961) who writes, "It (transfer of learning) refers to the extent to which knowledge and ability learned in one situation will be used in new and difficult situations". According to Heywood (1977) "the provision of learning experiences which place a student in the transfer situation is thus an important part of engineering education". This statement ofcourse is not exclusive to engineering education and is applicable to other disciplines as well.

Saupe (1961) has indicated several factors which help in the transfer of learning. Critical thinking is particularly important for transfer of learning. Following are the main abilities which, according to Saupe (1961) are useful for development of critical thinking and thereby transfer of learning.

1. Ability to recognize the existence of a problem,
2. Ability to define the problem,
3. Ability to select information pertinent to the problem,
4. Ability to recognize assumptions bearing on the problem,
5. Ability to make relevant hypotheses ,
6. Ability to draw conclusions validly from assumptions, hypothesis and pertinent information ,
7. Ability to judge the validity of process leading to the conclusion,
8. Ability to evaluate a conclusion in terms of its assessment.

As one closely analyses the above listed eight abilities considered so necessary to ensure the transfer of learning, it clearly emerges that these abilities ~~then seem to~~ cover a wide spectrum of professional traits like professional judgement, common sense, problem solving ability, initiative, decision-making ability, art of guestimation, leadership, intellectual ability, skills for data handling, etc; thus implying ~~thereby~~ that the educational objectives of the "transfer of learning" and the need for 'multiple objective assessment' of the student performance are indeed two sides of the same coin. Here, in this context, the term 'multiple objective' is taken to mean a set of objectives describing the total professional personality of the student. Therefore, these objectives, over and above the higher order objectives from the cognitive domain, also include certain qualities even from the affective domain.

Coming to the choice of teaching and learning techniques towards achieving educational objectives as stated under 'transfer of learning,' it will not be far from reality to say that the choice of the project method of learning is perhaps inherent in the very definition of the concept of "transfer of learning".

It is against the above frame of reference that one can then look at the PS model of higher education (as evolved at BITS) as one viable approach to introduce the transfer of learning and, hence, multiple objective assessment at the tertiary level. As explained in chapter 5, this model of the university system based on the PS theme, provides a student, after he has gone through the expositary method of class-room education, with an opportunity to involve himself in a project-method based learning experience which essentially places him in a transfer situation as represented by the professional environment.

The question under investigation in this chapter, however, intends to seek further integration of the learning experiences as perceived through the traditional class-room education with those as perceived through the theme of transfer of learning. In explicit terms then, the question posed is, is it is possible to introduce the rigour of transfer of learning into teaching of a course pursued under the class-room based traditional educational culture? and, if so, then how. Considering the fact that multiple-objective examination is the valid evaluation method for transfer of learning the question stated above then can be simplified as "how the class-room based examinations can be made multiple-objective in nature".

A case study of multiple objective examining in a class-room based course in engineering science is reported by Heywood (1977). In this case study, the examination structure developed had the following objectives:

(i) Knowledge and understanding

(ii) Comprehension

(iii) Problem finding

(iv) Application of principles to practice analysis

(v) Practical ability to conduct an investigation; practical ability to plan, implement and evaluate the solution to an engineering problem.

These objectives are multi-dimentional in nature and, as shown by Heywood (1977), can be achieved in class-room evaluation by making a project-work an integral part of the normally narrative method based classroom-method of teaching. However, what is most crucial in such an innovation is to have suitable real-life projects with which the teaching in class-room courses can be properly linked. This aspect has been particularly emphasized by Livingstone (1971) in the context of some management courses at Harvard Business School. According to Livingstone (1971) these courses, although project oriented, failed because they could not ensure the total gamut of the real-life problem-solving scenario starting from the stage of the problem identification to the stage of the problem completion.

Thus, what seems to emerge is the fact, that, while, by integrating with it the project-method-based teaching technique, the conventional class-room-based learning experience can be made a suitable candidate for imparting the educational objectives, characterised by the multiple-objective assessment, as perceived through the theme of transfer of learning, however for a reliable and valid outcome, it seems absolutely essential that the contents and culture of the project-method-based teaching scenario chosen for integration should faithfully simulate the stresses and strains peculiar to the syndrome of the real-life professional experience.

Indeed, the need for so strengthening the class-room based teaching and, hence, assessment techniques has always been argued by the educationists. In words of Heiss (1973), this would help student to develop the ability to organise, integrate and relate knowledge, to adopt study habits that will enhance learning throughout his lifetime and to take responsibility for his own education. Further, this would also assist the student in coordinating his outside reading with his course material, and test his ability to extemporise on problems and propositions. And, further, as pointed out by Heywood (1977), the incorporation of the multiple-objective assessment in the conventional class-room based approach to teaching would "provide an overall assessment of all facets of a student's learning" so that he can evaluate problems and propositions not previously met.

Interestingly, a need of the above sort has been intensely experienced at BITS in terms of the implementation of its recent educational innovation of the M.E.(Collaborative) programmes. More specifically, from the year 1979 onwards, as a natural corollary to its acceptance of the PS model of university education, the

Institute has initiated these full time Master of Engineering (Collaborative) programmes at its off-campus based educational centres in direct collaboration with professional organisations, which for the purpose of educational details pertaining to the programme requirements play the role of live laboratories.

These programmes are of two year duration, the first year consisting of eight courses, offered four per semester, and the second year consisting of internship courses. While the courses from the first year carry the elements of the syndrome of the campus-based class-rooms, the internship courses, consistent with their very nomenclature, are entirely constituted of experience based approach to teaching and learning processes evident in the professional world. Thus, it is during the first year of teaching that one comes across a situation where the traditional class-room based teaching methodology and the learning experience as under the transfer situation can co-exist; thus, in turn, creating the need for multiple-objective approach to student assessment.

Indeed, the scenario emerging here does not have one-to-one correspondence with that of the class-room based teaching experience as present in the conventional campus-based university model of higher education, but nevertheless, for the purpose of the analysis, it can certainly be taken to be a good approximation to the very conditions inherent in the educational innovation of integrating the theme of transfer of learning into the campus-based class-room approach to teaching and learning as present in the non-PS based, i.e. the traditional university model of higher education.

It is against the above frame of reference that the next section of this chapter then, abstracting the implementational details pertaining to the M.E.(Collaborative) programmes, presents a model of an examination system involving multiple-objective assessment for a class-room based approach to teaching characterised by the theme of transfer of learning.

7.3 The Model

As indicated earlier, this section presents a model of a multiple-objective examination system for class-room based course offerings characterized by the theme of 'transfer of learning'. Towards this, as a case-study, the section heavily draws upon the operational details pertaining to a typical M.E.(Collaborative) programme in the interdisciplinary and functional discipline of Project Engineering implemented by BITS at its off-campus educational centre at Calcutta.

As explained in the previous section, the M.E.(Collaborative) programme in the discipline of 'Project Engineering' is of two years duration and it is entirely implemented at the Institute's off-campus educational centre at Calcutta in collaboration with a well-known leading Indian engineering consultancy firm, namely, ~~the~~ Development Consultants Pvt. Ltd. (DCPL). The first year of the programme consists of eight courses, four each semester, while the second year consists entirely of only internship courses. The eight first year courses are, namely, Systems Engineering, Overview of Engineering, Design Methods & Decisions, Project Evaluation & Formulation, Social Engineering, Project Management, Plant Layout & Infrastructural Engineering, and Recent Advances in Technologies & Design Techniques. For further details in terms of the syllabus for these courses as also for other additional information on this particular or any other M.E. (Collaborative) programmes of the Institute, one may refer to the Institute's Bulletin, 1980-81.

For the purpose of the subject matter under study in this section, one may choose to further explore the educational organisation of the eight courses in the I year of the M.E. (Collaborative) programme under consideration. As can be seen from the syllabus for the above listed eight courses for the discipline of Project Engineering (BITS Bulletin 1980-81) the contents of these courses have essentially emerged as an abstraction of the principles, theories and techniques corresponding to the entire range of the functional activities that go to define the universe of a project engineer. These functional activities are: (i) Market study, (ii) Techno-Economic Feasibility Study, (iii) Financial & Investment Analysis, (iv) Basic Design Engineering, (v) Project Approvals - Administrative, Financial & Statutory, (vi) Site Development, (vii) Detailed Design Engineering,(viii) Project Scheduling & Cost Monitoring; Procurement Inspection & Expediting, Manufacturing, Construction and Erection, (ix) Commissioning, and (x) Production. Thus, the I year of the programme, essentially, seems to aim at providing a student with an analytical understanding as also an overall exposure of the entire range of the interdisciplinary and functional activities that go to define the area of project engineering.

It is in terms of the above stated educational objectives that one must then pursue the analysis of the teaching and assessment/evaluation techniques that can be adopted in conducting of the I year courses of the programme. In this context, for the purpose of convenience, one may concentrate on say, the I semester courses of the I year.

As indicated in the Institute's Bulletin of 1980-81 the four I semester courses from the I year of the M.E.(Collaborative) programme in Project Engineering are, namely, Systems Engineering (Syst Engg.), Overview of Engineering (OE), Design Methods & Decisions (DMD), and Project Evaluation & Formulation (PEF). In terms of their educational contents, broadly speaking, while the course 'Systems Engineering' (Syst.Engg.)introduces students to the advanced techniques for analysing the large and complex systems problems, the course 'Overview of Engineering'(OE), amongst other things, exposes students to the functional areas of market and techno-economic feasibility studies.

Against this, the course 'Design Methods & Decisions'(DMD) introduces the student to the system design situations in terms of the functional activities of basic and detailed design engineering, whereas the course "Project Evaluation & Formulation' (PEF) requires the student to directly plunge, as a member of, say, an engineering consultancy team, in a task pertaining to a preparation of a detailed project report (DPR) consistent with the apriori announced customer specifications.

Thus, in a totality, it seems to emerge that the I semester education from the I year of the programme under consideration amongst other things, essentially introduces the student to the entire gamut of the tasks pertaining to the preparation of a Detailed Project Report (DPR), with particular reference to the functional areas of Market Study, Feasibility Study, Site Development, Basic Design Engineering and Detailed Design Engineering*.

Indeed, it is to meet the demands of the type of the educational objectives as stated above that the teaching methodology adopted in the I year of the M.E.(Collaborative) programmes requires that, while the students are being introduced to the knowledge as also various theories and analytical techniques from the area of the subject matter under study, they must also be simultaneously exposed, consistent with the rigour of the professional culture, to the corresponding functional activities the sum total of which goes to define the discipline for which they (students) are being trained. Seen in an abstraction, it is this proposition that can be considered to be central to the choice of teaching and student assessment techniques (for the I year courses of the M.E.(Collaborative) programmes), representing a peculiar mix of the traditional teaching and assessment practices as followed in campus-based class-rooms and the learning experiences accompanied by a multiple-objective assessment, as provided through the theme of transfer of learning. In what

* It follows as a natural corollary to this that the II semester courses from the I year of the programme, amongst other things, then basically aim at exposing the student to the entire gamut of the tasks pertaining to the theme of implementation of a project (i.e. Project Management), with particular reference to functional areas of Project Engineering not covered in the I semester.

follows, this section essentially discusses the details of these teaching and, particularly, the student assessment techniques, taking the I semester courses of the I year of the M.E.(Collaborative programme in Project Engineering as a case study.

Table 7.1 presents unit-wise and weekly-hours-wise student-workload distribution for the I semester courses of the I year of the M.E.(Collaborative) programme in 'Project Engineering'. As can be seen from table 7.1, this first semester under consideration in all consists of 16 units, and, in terms of hours, the weekly-student-work-load is 53 hours, thus implying a work day of about 9 hours per day. These 53 hours are almost equally divided over the four courses of the I semester in the sense that, if the course 'Systems Engineering' covers 14 hours (i.e.26.5%) of the total weekly-student-work-load, each of the remaining three courses, namely, 'Overview of Engineering', 'Design Methods & Decisions', and 'Project Evaluation & Formulation' account for 13 hours (i.e. 24.5%) of the total weekly-student-work-load.

Against this, when viewed from the point of the learning-situation, the percentage distribution of the weekly-student-work-load turns out to be such that while the formal contact hours through lectures cover 9.4% of the total weekly load, the self study efforts to prepare oneself for attending lectures account for as much as 28.3% of the total weekly-student-work-load, and, while the formal contact in terms of special lectures, seminars, group-discussions, gap-lectures, preparation for site visits, etc. covers 5.7% of the weekly load, the self-study efforts in terms of preparation for participation in such type of

TABLE 7.1

Unit-wise and Weekly-Hours-Wise Student-Work-Load Distribution for the I semester Courses of the I year of the M.E.(Collaborative) Programme in "Project Engineering".

Course Title	Units L	P	T	U	Hours-wise Distribution of the Weekly Student Load: Weekly formal contact hrs. for Lectures(L)	Professional work weekly hrs.(P) as reflected through: (a)Professional involvement in various functional activities (b) Site visits (c)Preparation of detailed project reports towards stated professional goals, and (d)Home project on applications of modern techniques to real-life professional situations	Weekly formal contact hrs. in terms of special lectures, seminars, group discussions, gap-lectures case studies, preparation for site visits, etc. (T)	For student to prepare to attend Lectures: Self-study involves: a)Library Reading b)Revision of previous portion, c)Discussions with faculty & experts d)Home Assignments e)Home Projects	For student to prepare for participating in activities under Column 6: Self-study involves: a)Library Search b)Discussions with faculty and professional experts, c)Project work further this also includes weekly hrs. for site visits	Total weekly hours	Course-wise % Distribution of the weekly load
1	2				3	4	5	6	7	8	9
Systems Engineering	2	3	1	4	2	3	1	6	2	14	26.5%
Overview of Engineering	1	6	1	4	1	6	1	3	2	13	24.5%
Design Methods & Decisions	1	6	1	4	1	6	1	3	2	13	24.5%
Project Formulation & Evaluation	1	9	0	4	1	9	0	3	0	13	24.5%
Total				16	5	24	3	15	6	53	-
Learning-Situation-wise % distribution of the weekly hours					9.4%	45.3%	5.7%	28.3%	11.3%	-	100%

group-involvement based learning experiences account for 11.3% of the total weekly-student-work-load.

And, as a result then, the weekly professional work hours as reflected through professional involvement and other similar types of activities signifying the learning situations providing for transfer of learning, are found to cover as much as 45.3% of the total weekly-student-work-load.

Thus, when one analyses the contents of table 7.1, three main categories of the types of learning experiences as available through the I year courses of M.E.(Collaborative) programmes emerge. These three categories are, namely, (i) the traditional class-room based learning experience characterized by narrative approach to teaching as reflected through the lecture method, (ii) the group involvement based learning experiences, as reflected through seminars, group-discussions, home project/ assignments, case studies, common to both the university system as also the modern personnel training practices of the professional world, and (iii) transfer of learning based experiences as reflected through the student involvement in the ongoing functional activities of the profession or through the student participation in a project oriented activity consistent with the entire syndrome of the professional culture. Table 7.2 describes various techniques of teaching as also the evaluation components adopted for each of these three categories in terms of the educational organisation as evolved for the M.E. (Collaborative) programmes.

It is against the above frame of reference that table 7.2 gives a typical evaluation-component-wise distribution of marks

TABLE 7.2

Description of typical evaluation-components as also teaching techniques adapted in different learning experiences as available through I year courses of M.E. (Collaborative) Programmes

Type of learning experiences as available through the I year courses of the M.E.(Coll.) Programmes	Teaching Techniques Adapted	Evaluation Components Adapted
1. Traditional class-room based learning experience characterised by narrative approach to teaching	1. Lecture method	1. Tests 2. Comprehensive Examination
2. Group involvement based learning experiences common to both the university system as also the modern personnel training practices of the professional world. Here the evaluation objectives pursued are: (i) Improving of the knowledge of modern analytical techniques, (ii) widening of horizon and, (iii) improving of intercommunication skills	1. Seminar 2. Group Discussions 3. (Home) Projects/Assignments 4. Case Studies	1. Seminar 2. Group Discussions 3. Assessment of written reports 4. Viva
3. Learning experiences as provided through the transfer of learning	Project method based as reflected through: a) work oriented professional involvement in terms of various functional activities b) Site visits c) preparation of detailed project reports towards stated profe-ssional tasks. d) Home projects on applications of modern techniques of analysis to real-life professional situations.	1. Assessment of professional involvement and site visit components as done by the professional expert(s) under whom the student is attending these components. 2. Technical defence before a group consisting of faculty and professional experts. 3. Assessment of project reports by a team consisting of professional expert(s) and faculty. 4. Seminar

for the four I semester courses of the I year of the M.E. (Collaborative) programme in Project Engineering. An analysis of the contents of table 7.2 giving the course-wise distribution of evaluation weightages as corresponding to the different categories of learning experiences is then given in table 7.3. It is from this analysis one can see that, as the "professional involvement" based weekly-student-work-load increases from 3 hours for the course Systems Engineering to 6 hours for the courses OE and DMD to 9 hours for the course PEF, the evaluation weightage as assigned to the learning component under these courses corresponding to the traditional class-room based learning experience as reflected through the lecture method of teaching-decreases from 75% for Systems Engineering to 45% for OE and DMD to 35% for PEF. Against this, as one can see from table 7.4, for the above stated weekly professional involvement schedule, the evaluation weightage as assigned to the learning component under the theme of transfer of learning, as provided through the direct student involvement in the on-going functional activities or in the project-oriented tasks of real-life nature, increases from 25% for courses Systems Engineering and OE to 30% for the courses DMD to 65% for the course PEF.

It is in the context of the above details that one can now further explore the teaching organization for the I year courses of the M.E.(Collaborative) programmes as evolved by BITS.

Indeed, consistent with the educational organisation, each of the I year courses of the M.E.(Collaborative) programme in "Project Engineering" are taught by a team consisting of Institute faculty resident at the Institute's off-campus educational centre where the M.E.(Collaborative) programme is being conducted, and

TABLE 7.3

A Typical Evaluation-Component Weightage Distribution for the I Semester Courses of the I year of the M.E.(Coll.) Programme in Project Engineering

Systems Engineering — 2 3 1 4

Evaluation component	Maximum weightage
Test I	20
Test II	20
Comprehensive	35
Home Project(s) on application of Systems Engineering Techniques to real-life professional situations	15
Seminar	10
Total	100

Overview of Engineering — 1 6 1 4

Evaluation component	Maximum Weightage
Test I	10
Test II	10
Comprehensive	25
Professional involvement in Functional areas: (i) Market study, (ii) Techno-economic feasibility study	25
Home Project(s)*	15
Group Discussion*	15
Total	100

Design Methods & Decisions — 1 6 1 4

Evaluation Component	Maximum Weightage
Test I	10
Test II	10
Comprehensive	25
Professional involvement in terms of functional areas: (i) Introduction to site development (ii) Basic design engineering (iii) Detailed basic engineering	30
Home Project(s)*	15
Seminar*/Viva	10
Total	100

Project Evaluation and Formulation — 1 9 0 4

Evaluation Component	Maximum Weightage
Test I	10
Test II	10
Comprehensive	15
Professional involvement in terms of Detailed Project Report (DPR)	35
The written presentation of the Detailed Project Report	20
Defence of the Detailed Project Report	10
Total	100

* On topics aiming at the educational objectives of (i) improving of knowledge of modern analytical techniques, (ii) widening of horizon and (iii) improving of intercommunication skills.

TABLE 7.4

Course-wise distribution of Evaluation Weightages given to Different types of learning experiences as available during the I semester Courses of the I year of the M.E.(Collaborative) Programme in Project Engineering

Type of learning experiences as available through the I year courses of the M.E.(Coll.) Programme	Types of Evaluation Components used	Total Evaluation Weightages			
		Syst.Engg. 2 3 1 4	O.E. 1 6 1 4	D.M.D. 1 6 1 4	P.E.F. 1 9 0 4
1. Traditional class-room based learning experience characterised by narrative approach to learning	1. Lecture method 1. Tests 2. Comprehensive Examination	75	45	45	35
2. Group involvement based learning experiences common to both the university system as also the modern personnel training practices of the professional world. Here the evaluation objectives pursued are: (i) improving of the knowledge of modern analytical techniques, (ii) widening of horizon and, (iii) improving of intercommunication skills.	1. Seminar 2. Group Discussions 3. Assessment of written reports 4. Viva	-	30	25	-
3. Learning experiences as provided through the transfer of learning.	1. Assessment of professional involvement and site visit components as done by the professional expert(s) under whom the student is attending these components. 2. Technical defence before a group consisting of faculty and professional experts. 3. Assessment of project reports by a team consisting of professional experts and faculty. 4. Seminar	25	25	30	65
TOTAL WEIGHTAGE		100	100	100	100

tion. For each team one of the Institute faculty members from that team acts as an in-charge for that course. All these course-team-members together go to form a core team for the M.E. (Collaborative) programme. One of the Institute faculty members from this core team for the programme plays the role of the M.E. (Collaborative) Convenor for the programme under consideration.

Entire responsibility for organizing, teaching and evaluation operations pertaining to a course rests with the corresponding course team. More specifically, the task of delivering the "lecture component" of the teaching effort for any given course is entirely fulfilled by the course team members for that course. As regards to the group-involvement based teaching effort in terms of activities like seminars, group-discussions, case-studies, special lectures, etc., the same is also implemented by the course team members assisted by other members of the programme-core-team, but, wherever necessary, they also seek involvement of other professional experts in terms of the teaching as also the student assessment tasks.

However, when it comes to the teaching component corresponding to the theme of "transfer of learning", ~~here,~~ the professional experts from the collaborating organisation can be seen to play a pivotal role. More specifically, as explained in table 7.1, for the I semester courses of the I year of the M.E.(Collaborative) programme in Project Engineering as much as 45.3% of the weekly student load corresponds to the professional work based involvement in terms of activities such as:

(a) Professional involvement in various functional desks of the collaborating organisation,

(b) Site visits,

(c) Preparation of detailed project report towards a stated professional goal,

(d) home projects based on applications of modern analytical techniques of analysis to real-life professional situations, etc.

The above activities are pursued by the student outside formal contact hours in terms of lectures, seminars, group-discussions, and case studies.

In an operational terms, thus, in each of the semesters of the first year of the M.E.(Collaborative) programme, each student attends various functional desks of the collaborating organisation as per the apriori announced sequence and schedule of professional involvement. For each functional desk, the student reports to the apriori designated professional expert who is in-charge of the concerned functional desk. A similer operational pattern is also followed in case of the activity of "site visits".

One word about the educational technique adapted when a student attends the spectrum of the functional desks. By definition the student involvement at any given functional desk is in terms of the on-going professional tasks pursued by the collaborating organisation at that desk. Thus, the professional expert, while taking the responsibility of exposing the student to the details of the functional task under consideration, can also look upon him as a manpower input.

Coming to the student involvement in a "transfer of learning" based activity like the preparation of a detailed project report towards a stated professional goal, it may be mentioned that this kind of learning opportunity is given to the student, basically, under the course Project Evaluation and Formulation. Thus, in this course, the class may be divided into, say, two groups and each group may be given an engineering consultancy project as specified by the professional experts from the collaborating organization. Needless to say, these consultancy projects are drawn from the on-going professional activities pursued by the collaborating organisation. The task before each group then is to prepare its own detailed project report for the consultancy problem posed before it and, then, to technically defend the same before a team of examiners as drawn from the professional experts involved in specifying the project as also the members of the course-team and the programme-core-team.

Finally, it may be mentioned that, during the I semester of the I year of the M.E.(Collaborative) programme in Project Engineering, it is the course "Systems Engineering" in which students are given home project(s) in terms of studying application of systems techniques to real-life professional situations as identified through the on-going functional activities of the collaborating organization. Thus, as indicated in table 7.2, the learning effort as reflected through these home projects also comes under the theme of transfer of learning. As a result, the course-team, assisted by the other members of the programme-core-team, also seeks participation of appropriate professional experts towards guiding and examining students in terms of their involvement in such the home projects.

The above then is a brief description of the organisation of the teaching effort for the I year courses of the M.E. (Collaborative) programmes as evolved by BITS, Pilani. It is based on these details that one can now describe the model for the evaluation system as is emerging for the above mentioned first year courses.

Table 7.2 has listed various evaluation components that are adapted for the types of the teaching efforts described above. Thus, the lecture method based teaching component of any given I year M.E.(Collaborative) course is, normally, evaluated through tests and comprehensives. Understandably, these tests and comprehensive examinations are heavily biased towards the lower order skills from the cognitive domain, with greater emphasis on 'knowledge'. As a result, these tests and comprehensives are designed, administered and assessed by the corresponding course-team members in a similar manner as that prevalent in the campus based class-rooms of the traditional university system.

Against this, as described earlier, the group involvement based teaching component of any given course is assessed through the evaluation components of seminar, group-discussions, written reports and viva, and in this assessment the professional experts from the collaborating organisation also participate along with the course team as well as the programme-core-team members, bringing the evaluation components in terms of their structure and objectives closer to the practices followed in the modern professional set-ups. Thus, the evaluation certainly then aims at higher order objectives, wherever feasible going even beyond the cognitive domain. As a result, the design of these evaluation components follows same structure as the one presented through the evaluation-component-vs-educational-objective-based specification matrix for PS-II course presented in Appendix C.

Finally, as indicated in table 7.2 and explained subsequently, the "transfer of learning" oriented professional-work-based teaching component of any given I year course has in its implementation a central role played by the appropriately identified professional experts of the collaborating organisation in terms of the tasks of student teaching and evaluation. Indeed, as discussed earlier, it is such an involvement on the part of the professional experts that goes a long way towards making it feasible to integrate right from the start of the student educational process under the M.E.(Collaborative) programme with the on-going production processes of the participating organisations. And, this in turn then goes to ensure almost a total equivalence between the student assessment practices followed for the teaching component under consideration and those prevalent in the professional world. Thus, understandably, the evaluation components for this teaching component centred around the theme of "transfer of learning" are multiple-objective in nature.

As has been described earlier, the course "OE" has the professional-work-based teaching component implemented through the student professional involvement in functional areas of

(i) Market Study and (ii) Feasibility study, while the course "DMD" has its professional-work-based teaching component implemented through the student professional involvement in functional areas of (iii) Introduction to Site Development, (iv) Basic Design Engineering and (v) Detailed Design Engineering. Tables 7.5 through 7.9 describe the evaluation components for the above mentioned five functional desks, giving for each evaluation component the entire spectrum of objectives assessed along with their weightages.

Similarly, the site-visit type professional-work-based teaching component is also evaluated consistent with the personnel assessment practices existing in the professional world. Finally, as regards to the tasks such as preparation of the detailed project reports or work in terms of home projects etc., as coming under the theme of professional-work-based teaching component, it may be mentioned that these (i.e. tasks) are also evaluated through evaluation components like written project reports, viva etc. which are essentially multiple-objective in nature. The design of these evaluation components, understandably, follows the same structure as the one presented through the evaluation-component-vs-educational-objective-based specification matrix for the PS-II course given in Appendix C.

Thus, with reference to the I semester courses of the I year of the M.E.(Collaborative)programme in Project Engineering, the professional-involvement-based evaluation components as presented through tables 7.5 to 7.9, and the evaluation components of viva, seminar, group-discussions and project-report as described in the evaluation-component-vs-educational-objective

TABLE 7.5

A Typical Evaluation Scheme for the Professional Involvement in the Functional Area of "Market Study" under the I semester course "Overview of Engineering" From the I year of the M.E.(Collaborative) Programme in "Project Engineering" As Evolved by BITS at its Off-campus Educational Centre at Calcutta in Collaboration with DCPL

1. Name of the Student ________________ 2. ID No. ________________ 3. Semester ______________
4. Name & Designation of the Professional Expert from the Collaborating Organisation Under whom the Functional Activity is Implemented ____________________________.

Criterion of evaluation / Activity Break-up	Understanding of Concepts	Application of Knowledge	Computational Skill	Presentation Ability	Accuracy in Work	Speed of Work	Ability for Hard Work	Sense of Responsibility	Decision Making	Maximum Marks	Marks Obtained
Data Collection										20	
Data analysis & Forecasting										40	
Application of Techniques										40	
										100	

Signature of the Professional Expert

TABLE 7.6

A Typical Evaluation Scheme for the Professional Involvement in the Functional Area of "Feasibility Study" under the I Semester Course "Overview of Engineering" From the I Year of the M.E.(Collaborative) Programme in "Project Engineering" as Evolved by BITS at its Off-campus Educational Centre at Calcutta in Collaboration with DCPL

1. Name of the student ______________ 2. ID No. ______________ 3. Semester __________

4. Name & Designation of the Professional Expert from the Collaborating Organisation Under Whom the Functional Activity is Implemented ____________________.

Criterion for evaluation / Break-up of activity	Understanding of Concepts	Application of Knowledge	Computational Skill	Ability of Presentation	Accuracy of Work	Speed of Work	Ability for Hard Work	Sense of Responsibility	Decision Making	Maximum Marks	Marks Obtained
Choice of Plant Location & Size										30	
Choice of technology										30	
Cost estimation										40	
Feasibility Analysis										25	
Financial Projection and Analysis										25	
										150	

Signature of the Professional Expert

TABLE 7.7

A Typical Evaluation Scheme for the Professional Involvement in the Functional Area of "Introduction to Site Development" under the I Semester Course "Design Methods & Decisions" From the I year of the M.E.(Collaborative) Programme in "Project Engineering" as Evolved by BITS at its Off-campus Educational Centre at Calcutta in Collaboration with DCPL

1. Name of the student ______________________ 2. ID No. ______________ 3. Semester __________

4. Name & Designation of the Professional Expert from the Collaborating Organisation Under Whom the Functional Activity is Implemented ______________________.

Break-up of activity \ Criterion for evaluation	Understanding of Concepts	Application of Knowledge	Skill in Work	Ability for Presentation	Accuracy in Work	Speed in Work	Ability for Hardwork	Sense of Responsibility	Decision Making	Maximum Marks	Marks Obtained
Site investigation										15	
Site selection										15	
Site development										10	
										40	

Signature of the Professional Expert

TABLE 7.8

A Typical Evaluation Scheme for the Professional Involvement in the Functional Area of "Basic Design Engineering" under the I Semester Course "Design Methods & Decisions" from the I year of the M.E.(Collaborative) Programme in "Project Engineering" as Evolved by BITS at its Off-campus Educational Centre at Calcutta in Collaboration with DCPL

1. Name of the student ____________ 2. ID No. ____________ 3. Semester ____________

4. Name & Designation of the Professional Expert from the Collaborating Organisation Under Whom the Functional Activity is Implemented: ____________.

Break-up of activity \ Criterion for evaluation	Knowledge of fundamentals of his discipline	Knowledge of fundamentals of other disciplines	Professional understanding of activity	Computational Skill	Drawing Skill	Ability for technical presentation	Accuracy in work	Speed of work	Ability for hard work	Team Spirit	Sense of responsibility	Decision making	Maximum Marks	Marks Obtained
Flow-sheet preparation													12	
Choice of the process													12	
Major equipment selection													12	
Space requirements													12	
Plant layout													12	
Utility & services layout													12	
Ancillaries layout													11	
Plot plan													11	
Electric power requirements													12	
Equipment protection													12	
Electrical earthing													12	
													130	

Signature of the Professional Expert

TABLE 7.9

A Typical Evaluation Scheme for the Professional Involvement in the Functional Area of "Detailed Design Engineering" under the I Semester Course "Design Methods & Decisions" from the I year of the M.E.(Collaborative) Programme in "Project Engineering" as evolved by BITS at its Off-campus Educational Centre at Calcutta in Collaboration with DCPL

1. Name of the student ________________ 2. ID No. ______________ 3. Semester ______________

4. Name & Designation of the Professional Expert from the Collaborating Organisation Under Whom the Functional Activity is Implemented: ______________________________.

Break-up of activity / Criterion for evaluation	Knowledge of fundamentals in his discipline	Knowledge of other disciplines	Understanding of Concepts	Computational Skill	Drawing Skill	Technical Presentation	Accuracy in Work	Speed of work	Team Spirit	Sense of responsibility	Decision Making	Maximum Marks	Marks Obtained
Flow-sheet finalization												20	
Detailed layout												20	
Specification preparation												10	
Design of supporting systems like foundations, piping, etc.												40	
Procurement												20	
Equipment design												20	
												130	

Signature of the Professional Expert

based specification matrix presented in Appendix C , along with the traditional evaluation components of tests and comprehensive examinations, go to describe a typical model of a multiple-objective based examination system for a class-room type teaching effort.

As this stage it would be worth considering the issues that would be emerging if the above model were to be implemented in a campus-based class-room. From the details of the model it follows that an evaluation and, therefore, teaching component, as reflected through the 'professional involvement', is central to a multi-objective examination system, if the same (i.e. the multiple-objective examination system) were to be implemented in a class-room-based teaching effort. In such a case, the question that will need to be answered, if the multiple-objective assessment were to be incorporated in a campus-based classroom, would be what can then be the analogue of the 'professional involvement' in a campus-based environment? A natural response to this query could be in terms of the theme of the sponsored-research-based consultancy that an university can undertake.

If the above scenario were to be feasible, then certainly the university faculty, constituting the members of the sponsored research-based-consultancy team, can play the role of the immediate 'research professionals', under whom the students of the campus-based course under consideration can attend the project-work-based learning experience characterised by the theme of 'transfer of learning'. In such case, then, the very demands of the sponsored research activity would make it possible to evaluate a student in terms of his multi-dimensional professional personality.

However, to begin with, it is important to realize that the task of orienting the traditional research base of the university system so as to accommodate the demands of the professional culture as demonstrated by the sponsored-research-based-consultancy-work is not an easy one. And, the task of integrating the traditional teaching role of a university with its research role visualised in an environmental context is further difficult.

Seen in a broader perspective, the tasks indicated above are the tasks of the educational and research organisation, without which it would indeed be very difficult to introduce the multiple-objective assessment in a campus based classroom. Purusal of these tasks would go to mean new attitudes to teaching and research and this in the context of the system of higher education can mean a host of things starting from curricula re-organisation to innovative approach to the funding of research at the universities.

Indeed above issues, however interesting, are beyond the scope of this thesis. Thus, what then clearly emerges is the observation that the model of the multiple-objective evaluation system as developed in this chapter can be introduced in a campus-based class-room, given certain basic attitudes are accepted by the academic community, firstly, in terms of transforming the traditional research base into that of the sponsored -research-cum-consultancy, and, secondly, in terms of integrating the teaching activity with the demands of the above type of a research-cum-consultancy based evolved in an environmental context.

7.4 Conclusion

Thus, this chapter has basically dealt with the question of evolving of a model for the multiple objective evaluation system for a class-room based teaching effort. Towards this the chapter has heavily drawn from the educational experience gained by BITS in terms of its recently introduced M.E. (Collaborative) programmes. Finally, the chapter has also discussed the issues of educational and research organisation that should be emerging if the above model of the multiple-objective education system were to be incorporated in the campus-based classroom teaching effort.

CHAPTER 8

CONCLUSION

This Thesis has studied some aspects of the internal continuous evaluation system at the tertiary level. Towards this, the thesis has extensively drawn from the data-base as available through the on-going educational operations at BITS, Pilani. Below are summarized some of the main observations or conclusions emerging from this study.

Various facets of the internal evaluation system investigated in this thesis include the study of the teachers' perception of the objectives of education as also of the objectives, techniques and nature of examinations. This investigation was carried out with the help of the survey method by administering questionnaires to the Institute faculty. Thus, while 51 faculty members drawn from the entire spectrum of engineering, science and humanities disciplines answered the questionnaire on the objectives of education, as many as 72 responded to the questionnaire on the objectives, techniques and nature of examinations.

From the analysis of the above investigation it emerged that, on the whole, the teachers perceive the qualities or educational objectives that the education should pursue quite in isolation from the qualities that are needed in life. In specific terms the teacher responses to the question 'what qualities life needs most' showed poor correlation with the teachers responses to the question 'what education should achieve'. Further, the teacher responses to the question 'what qualities life needs most' also showed poor correlation with their responses to the questions 'what qualities education can achieve' and 'what qualities education has achieved'. However, the teacher responses to the questions 'what qualities education should achieve', 'what qualities education can achieve'and 'what qualities education has achieved' showed high to moderate correlations between themselves.

Coming to the objectives of examinations, as far as the evaluation components of quizzes, tests and comprehensives are concerned, it was observed that they are normally designed around the lower order skills as represented through the cognitive domain of objectives. Further, the investigation also noted that the majority of the teacher sample studied thought that the higher order skills like leadership, co-operation, sense of responsibility, decision-making, etc. could be assessed through classroom based evaluation. Towards this, the evaluation components of home-projects, home-assignments and project-work were observed to be most popular with teachers.

As regards the techniques of examinations, it emerged that the majority of teachers from the sample studied incorporated one or other evaluation component in addition to those of quizzes, tests and comprehensives. The most popular evaluation component in this context was observed to be that of the project work, followed by the evaluation components of home-assignments and lab-work and seminar.

Coming to the nature of examinations, it was observed that it is 'all or mostly closed-book' type examinations that constitute between 66-72% of the total sample sizes studied in this context for the evaluation components of quizzes, tests and comprehensives; the remaining percentage being accounted for by the 'mostly open-book' type examinations along with the 'mixed' type examinations. Further, it also emerged that while quizzes consist mainly of objective and short-answer type questions, the tests and comprehensives include mainly short answer and long answer type questions.

Next, the thesis has concerned itself with the reliability analysis for the examinations under the internal continuous evaluation system using the AOV approach. Towards this, the thesis in detail analysed 19 selected courses from the semester-wise offerings at BITS. Understandably, these courses were drawn from the entire spectrum of engineering, science and humanities disciplines and represented all the levels of the system of higher education. Thus, in this context, the thesis studied in all 83 examinations, consisting of 7 quizzes, 52 tests and 24 comprehensives. In specific terms, the reliability co-efficients for the above examinations were analysed in terms of various parameters such as the number of questions in an examination, number of students in an examination, duration of the examination, level of the course, discipline of the course, type of the examination (i.e. long-answer, short-answer,etc.), category of the courses in terms of core course or professional course and, finally, the aspect of choice in an examination.

Thus, it was observed that larger the number of questions in an examination, greater the possibility of its reliability being higher. The examination reliability was also found to get influenced by the number of students in an examination in a manner that larger the number, greater the possibility of the reliability being more. But then, interestingly, it was also observed that invariably the examinations with larger number of students were also characterized by larger number of questions. Coming to the examination duration, it was noted that normally the examinations with longer durations demonstrated higher reliability. Further, such examinations were also found to have larger number of questions.

As regards the parameter of the level of the course, it was observed that, normally the average reliability of an examination decreased as one moved from courses at the I and II year levels to the courses at the IV and V year levels. Here also it was noted that, on an average, the examinations for the courses at the lower level had more number of questions than that for the examinations for the courses at the higher level. Coming to the discipline of the course, for the sample studied, the examinations for the courses from science, humanities and management disciplines demonstrated better average reliability as compared to that for the examinations for the courses from the engineering discipline[5]. In this context, also, it may be mentioned that, normally, the examinations from science and management courses were found to incorporate more number of questions than that for the examinations from the humanities and engineering disciplines, in that order.

Coming to the parameter of the type of examination, it was observed that examinations with short answer type questions normally demonstrated better average reliability than that for the examinations with the long answer type questions. In this context, too, it was noted that, on an average, examinations with short answer type questions are also characterized by a larger number of questions. As regards to the category of the course, it was noted that the examinations from the core level courses demonstrated higher average reliability than that for the examinations for the professional level courses. Here, too, it was observed that, normally, the examinations for the core courses are characterized by higher number of questions than that for the examinations for the professional level courses. Finally, as the sample ~~of~~ the number ~~of~~ for examinations with choice was too small, no specific relationship emerged vis-a-vis the reliability and the choice in the examination except the observation that the faculty, on the whole, does not seem to encourage the practice of giving choice in the examination.

Thus, seen from every possible angle, it emerged that, on the whole, higher the number of questions in an examination, greater the possibility of its reliability being more.

Further, in terms of the data-base as available from the selected courses, the thesis also studied 303 evaluation-component-pairs for their inter-correlations. These evaluation components consisted of a wide spectrum of types of examinations such as class-room quizzes, laboratory quizzes, laboratory-work-projects, class-room project reports, home-assignments, seminars, tests, comprehensives, etc.

From the above investigation, it was observed that over 98% of the evaluation-component-pairs demonstrated positive correlations. In specific terms, as many as 81.85% of the total evaluation-component-pairs studied had correlations between the levels positive moderate to positive low. Thus, the percentage of the evaluation-component-pairs having positive moderate correlations was found to be 48.85%.

Finally, consistent with the constraint on the availability of the data, as many as 17 courses out of the 19 selected were studied for the course reliability. All these courses except one, demonstrated course reliabilities over 0.7, the course reliability for the remaining course being as high as 0.66. Thus, on the whole, the courses were observed to have satisfactory reliability co-efficients.

Another important aspect that the thesis has pursued in the context of the study of the internal continuous evaluation system pertains to the validity of examinations with specific reference to the criterion-related validity. For this purpose, the thesis selected the samples as emerging from the 1976 input to BITS of 310 students and 1977 input of 316 students. In specific terms, using the examination records for the above batches as the data-base, the thesis investigated the predictive validity for the student performance in 'higher secondary' as also for the student performances in the courses of 'Concepts in Science' and 'Modern Physics', the various other appropriately selected student performance measuring indices as available through the internal continuous evaluation system as at BITS playing the role of criteria.

Thus, assuming a temporal invariance in the character of the two batches of 1976 and 1977, it was observed that on the whole, whatever may , normally, be the criterion, the performance in Modern Physics showed itself to be a far better predictor than the performance in Concepts in Science and also the Higher Secondary performance; and, when forced to discriminate further, the performance in Concepts in Science, though by itself a poor predictor, was found to be a slightly better predictor than the performance in Higher Secondary.

Further, while on the whole the Modern Physics performance was found to be a better predictor, within the groups of disciplines it was noted to be more suitable as a predictor for Science group of students than for the students from the Applied Science & Technique Oriented disciplines, than for Engineering group of students, than for the Management group of students, in than order. And, within the Science group, the Modern Physics performance as a predictor was observed to be more suitable for disciplines of Mathematics and Physics than for the disciplines of Biology and Chemistry.

Against the above, for the engineering group of disciplines, the Modern Physics performance was found to be more suitable as predictor for the discipline of Electrical & Electronics Engineering (EEE), than for the discipline of Civil Engineering, than for the discipline of Chemical

Engineering, than for the discipline of Mechanical Engineering, in that order. As regards to the group of Applied Science & Technique Oriented disciplines, the Modern Physics as predictor was observed to be more suitable for the students from the Computer Science discipline, than for the students from discipline of Instrumentation, than for the students from Pharmacy discipline, in that order.

Further, grade point average (GPA) in Physics courses and CGPA (Cumulative Grade Point Average in each discipline) emerged as the best criteria against the Modern Physics as predictor for each of the discipline or the groups of disciplines scanning the entire spectrum from 'Sciences' to 'Engineering' to 'Applied Science & Technique Oriented disciplines'.

Finally, coming to the study of the concurrent validity, it was observed that for all samples, except in the case of the sample for mechanical engineering students, the correlations between the performances pertaining to the GPA in ENGG courses and GPA in their own discipline courses as also the correlations between the performances pertaining to the GPA in ENGG courses and CGPA in their individual programmes showed correlations levels between moderate to high.

Thus, in summary, through the above stated validity analysis the thesis has evolved a detailed methodology for the criterion-related-validity study under the internal continuous evaluation system at the tertiary level.

Apart from the reliability and the validity analysis for the campus based class-room examinations, the thesis also investigated the evaluation system for the Practice School(PS) system of education. As BITS, Pilani has implemented this scheme of education leading to the institutionalization of the linkages between the professional world and the academic world, naturally, its (Institute's) corresponding educational operations provided a meaningful data-base for such a study. In specific terms, the evaluation data as available from the two PS-II sessions at Nagda PS centre for the I semester 1976-77 and the I semester 1977-78 was used in the above context.

Thus, it was observed that the PS evaluation, which is multiple-objective in its character, is a peculiar combination of the traditional evaluation components as adapted in the campus-based university system and the on-going personnel appraisal practices in the industrial/ professional set-ups. Further, from the analysis of the PS evaluation data-base as available from the Nagda PS centre, it was noted that the evaluation component of quiz, as implemented under the PS surroundings, demonstrated lower reliability than that for the evaluation component of viva when similarly conducted under the PS set-up.

Further, within the framework of the sample studied, it was observed that the PS evaluation components of seminar, group-discussions and project-report demonstrated a fairly satisfactory examiner reliability. However, the PS evaluation component of 'observation' was found to have a very poor examiner reliability.

Coming to the study of the course reliability for the PS, both the PS-II sessions demonstrated high course reliability. Further, for the samples studied, all the inter-correlations between various PS evaluation components were observed tobe positive. More specifically, the evaluation-component pair of the seminar and the group-discussion demonstrated the best correlation, while, on the whole, the evaluation components of the 'diary' showed very poor correlations with all other components of evaluation. Finally, the thesis, also investigated content validity of the PS evaluation, which was observed to be satisfactory.

In the end, as a natural corollary to the analysis of the PS evaluation, the thesis studied the question as to how the multiple-objective assessment characterized by the higher order skills can be implemented in a class-room based situations. Towards this, thesis has devolved a model of the class-room based multiple-objective evaluation based on the theme of 'transfer of learning'.

In specific terms, this model building excercise has heavily drawn from the educational-organisation-based concepts corresponding to the Master of Engineering (Collaborative) programmes, that follow as a natural outcome of the PS based organisation of a university system.

The above then, is a brief summary of the conclusions of the various investigations as carried through this thesis. Some interesting issues emerge from the studies conducted under this thesis which could be further investigated in the futuristic context. For example, it will be interesting to determine the educational objectives that a system of higher education may be expected to pursue consistent with its social co-ordinates. Similarly, for examinations that showed poor inter-correlations, it should be worthwhile to investigate the factors behind the same so that the necessary improvements in the context can be brought out.

Another aspect that can be pursued in the futuristic context can be the content validation studies for the campus based class-room courses. Further, the question of the examiner reliability can also be studied for the campus-based class-room courses. As regards to the criterion-related validity studies, using the technique presented in this thesis, the same can now be carried out for a large number of examinations under different courses. Coming to the PS evaluation, reliability and the validity studies for the same, using the techniques developed in the thesis, can be conducted, for a larger sample so as to be able to seek necessary generalisations. It goes without saying that such investigations of the examinations would in the end analysis go a long way towards strengthening the system performance and thus in turn bring out further strengthening of educational details.

Further, it would also be interesting to study as to what campus-based education-cum-research-organisation can make it possible to bring in the multiple-objective assessment in the campus-based class-room situations. Further, it should also be worthwhile to apply the reliability techniques to the examinations as coming under the courses for the M.E. (Collaborative) programmes.

Finally, for the above to be feasible, it will be extremely important to initiate a kind of an educational research cell within each of the universities. In spite of the acceptance of the importance of such kind of studies, unfortunately, much is still desired in the above context even in case of the best of the universities. This, then, is a case for one to start such research efforts, as it is through this alone that the much needed improvements can come in the examinations at the tertiary level, which in turn would strengthen the design as also the delivery of the contents of the courses, and hence programmes, as under the system of higher education.

REFERENCES

CHAPTER - 1

Calcutta University Commission Report (1917-19) Calcutta Superintendent Government Printing, India. 1919.

Examination Reform - A Plan of Action (The Recommendations of the Zonal Workshops), (1976). University Grants Commission.

Gayen, A.K. (1970). Examination and Evaluation in Higher Education. Paper Presented at the tenth quinquennial Conference of Inter-University Board of India and Ceylon, Madurai.

Gayen, A.K., Nanda, P.B., Mathur, R.K., Duari, P., Dubey S.D. & Bhattacharya, N. (1961). Measurement of Achievement in Mathematics. Kharagpur: Indian Institute of Technology, 1961.

Gayen, A.K., Nanda, P.B., Duari, P., Bhattacharya, N., Mukherjee, H.N., & Mathur, R.K. (1962). Measurement of Achievement in English. Kharagpur, IIT.

Harper, A.E., Jr. and Misra, V.S. (1976). Research on Examinations in India. NCERT, New Delhi.

Indian Education Commission Report (1881-82).

Indian University Commission Report (1902).

Inter-University Board of India & Ceylon (1924). Examination in Higher Education. New Delhi, 1972.

Kahlon, A.S. (1977). Students' Evaluation and Internal Assessment, University News, 15 (3). Feb. 1, 1977. pp. 56,57.

Kothari Commission Report (1964-66) -
See Report of the Education Commission (1964-66)

Lele, P.T., Patel, P.M., Parikh, N.P., & Palkar, S.G. (1962). Empirical Studies in Examinations. Baroda: Examination Reform and Research Unit, M.S. University Baroda, March, 1962.

Lingamurthy, V. (1970). Examination Reform - Problems of Internal Assessment. Educational India, 36(7), Jan. 70. pp. 231-234 & 251.

Ministry of Education Report (1956)

Misra, V.S. (1973). Proposal for Examination Reform. Christian Education, 52(4), December 1973: 8-21.

Mitra, C.R. (1973). Educational Innovation in Britain and Common Interest with India. Published in New Frontiers in Education, Delhi. Volume III, No.3. December 1973.

Natarajan, V. (1977). Monograph on Internal Assessment for Universities. A.I.U. Publication.

Radhakrishnan Commission Report (1952) - See Secondary Education Commission

Raina, T.N. (1967). Relâtionship between External Examination and Internal Assessment. Teacher Education, July, 1967.

Rao, T.S.M. (1970). Examination System in India. Need for Fresh Approach. Paper presented at the tenth quinquennial Conference of Inter-University Board of India and Ceylon, Madurai

Report of the Education Commission (1964-66). Ministry of Education, Government of India, Published by Manager of Publication, Delhi - 1966.

Sargent Report (1944). Report of the Central Advisory Board of Education on Post-War Educational Development in India.

SCERT (1976). Scheme of Comprehensive Internal Assessment and the Manual of Instructions, Parts I & II. Published by Directorate of Government Examinations and SCERT, Tamil Nadu.

Secondary Education Commission Report (1952) -
See also Radhakrishnan Commission

Secondary Education Committee Assam Report (1965).

Singh, D.P. (1970). Reform in Examination System. Paper presented at the tenth quinquennial Conference of Inter-University Board of India and Ceylon, Madurai.

Singh, R.M. (1971). Examination Reform: Internal Assessment. Educational India. 38(3); Sept., 1971: 79-81.

Singha, H.S. (1977). Undergraduate Examinations - Procedure and Problems. The Management of Examinations, AIU, New Delhi.

Taylor, H.J. (1962). The Combination of Internal Assessment with Examination Results. Report on Examination Reform. UGC

The Report of the University Education Commission (1948-49). Published by the Manager of Publication, Govt. of India, Delhi 1951.

Wood's Education Despatch Report (1854).

CHAPTER - 2

Beard, R.M., Healy, F.G. and Holloway P.J. (1974). Objectives in Higher Education, Society for Research into Higher Education, London.

Bloom, B.S. (Ed.) (1979a). Taxanomy of Educational Objectives. Volume 1: Cognitive Domain, Longmans Green London.

Bloom, B.S. (Ed.) (1979b). Taxonomy of Educational Objectives. Volume 2: Affective Domain, Longmans Green London.

Burke, R.J. (1969). Some Preliminary Data on the use of Self-Evaluations and Peer Ratings in Assigning University Courses Grades. Journal of Educational Research.

Carter, G. and Lee L.S. (1974). University 1st year Electrical Engineering Examinations, International Journal of Electrical Engineering Education.

Cowan, A. (1971). Survey (International) of Dental Undergraduate Curricula. Journal of Irish Dental Association.

Dew-Hughes, D.; Heywood, J; Mortimor, E.G.; Montagu-Pollock H.; Oldfield, F., (1966). A Group Study of Staff and Student Attitudes to the Distant Minor. A Report to the Leverhulme Foundation on Research into Higher Education.

Dressel, P.L. (1971). Values, Cognitive and Affective. Journal of Higher Education, 42, (5). 400.

Eriksson, B.O. (1968). Planning and Construction of a University Course in Educational Technology. Institute of Education, Gotenburg University, Sweden.

Freeman, J. and Byrne, P.S. (1973). The Assessment of Post-graduate Training in General Practice. Society for Research into Higher Education, London.

Heywood, J. (1977). Assessment in Higher Education. John Wiley & Sons, London, New York, Sydney, Toronto.

Hull, F.W.(Ed.) (1971) Restructing Science Education for the Undergraduate Liberal Arts College and University, Center for the Study of Higher Education, University of Toledo, Ohio.

Kaysen, C. (1973). Content and Context: Essays on College Education, Mc-Graw Hill, New York. (Includes essays on the humanities, social sciences, art, science and liberal education).

Lee, L.S. (1969). Towards a Classification of the Objectives of Undergraduate, Practice Work in Mechanical Engineering. Thesis. University of Lancaster Library.

Long, R.S., Meltzer, N.S. and P.J. Hilton (1970). Research in Mathematics Education. Educational Studies in Mathematics,

Mac Ewan, M. (1970). New Directions in Architectural Education, RIBA Journal.

Mc Guire C.H. (1967). An Evaluation Model for Professional Education, College of Medicine, University of Illinois, Chicago.

Miller, G.E. (1962). Teaching and Learning in Medical School, Harvard U.P., Cambridge, Mass.

Perren, G. and Trim, J.L.M. (Eds.) (1971). Applications of Linguistics, University Press, Cambridge.

Plumb, J.H. (Ed.) (1964) Crisis in the Humanities, Pelican, London. (Classics, Philosophy, Sociology, History, Fine Arts, Economics, Literature).

Russel, G.F.M. and Walton H.J. (Eds.) (1970). The Training of Psychiatrists: Proceedings of the Conference on Post-graduate Psychiatric Education, March, 1969.

Sheffield, E.F. (Ed.) (1970). Curriculum Innovations in Arts and Science. Report of A Canadian University Workshop, The University, Toronto.

Tyler, R.W. (1949). Achievement Testing and Curriculum Construction. In E.G. Williamson (Ed.) Trends in Student Personnel Work, University of Minnesota, Minneapolis.

UNESCO (1966). A Survey of Teaching Physics at Universities. (Czechoslovakia, W. Germany, France, USSR, UK and USA), UNESCO, Paris.

Vaughan, J.C. and Rouve S. (1973). Survey of Curricula and Performance in Modern Languages. Centre for Information on Language Teaching and Research, London.

Wain, G.T. (1972). Mathematics in Colleges of Education (U.K.). Int. J. Maths. Education in Science and Technology, 3, 191.

World Health Organisation (1974). Joint ILO/WHO Conference on the Teaching of Occupational Health and Safety, WHO, Copenhagen.

Yudkin, M. (Ed.) (1971). General Education, Penguin, London. (Role of History, Sociology, English, Maths, Experimental Science and Philosophy in General Education).

CHAPTER - 3

Baruah R.K. (1975). A Hand book of Scaling. Examination Research Unit, Gauhati University

Dave, R.H. (1971). Reforming Written Examination: Some Measures. Education Quarterly, 23 (2).

Ebel, R.L. (1966). Measuring Educational Achievement. Prentice-Hall of India Pvt. Ltd., New Delhi.

Ebel, R.L. (1979). Measuring Educational Achievement, Third Edition. Prentice-Hall INC, Englewood Cliffs, New Jersey 07632.

Gayen, A.K., Nanda, P.B., Mathur R.K., Duari, P., Dubey, S.D. & Bhattacharya N. (1961). Measurement of Achievement in Mathematics, IIT, Kharagpur.

Gayen, A.K., Nanda, P.B., Duari, P., Bhattacharya, N., Mukherjee, H.N., & Mathur, R.K. (1962). Measurement of Achievement in English, IIT, Kharagpur.

Gayen, A.K. (1970). Examination and Evaluation in Higher Education Paper presented at the Tenth Quinquennial Conference of Inter-university Board of India and Ceylon, Madurai.

George, A. (1964). A Study on Assessment of Standard of Valuation of Comparative Performance of Colleges in the Pre-University Examination. Kerala: Department of Statistics, Kerala University.

Guilford, J.P. (1978), Second Edition. Second Reprint. Tata McGraw Hill Publication

Harper, A.E., Jr. (1960), Recent Advances in Psychometry. Silver Jubilee Volume of Vidya Bhawan Society, Masulipatnam.

Harper, A.E. Jr. (1962). Chance in the Traditional Examination. Journal of Education and Psychology. 20, 100-118

Harper, A.E. Jr. (1963a). Marking Objective Examinations; Teacher Education, 8, 10-17.

Harper, A.E. Jr. (1963b). Objective and Traditional Examinations: Some Research, Education and Psychology Review, Spring 1963, 63-78.

Harper, A.E. Jr. (1966a). Examiner Reliability. A Study of Ninety Experienced Examiners Marking the Same Ten Answer Books. Unpublished MSS

Harper, A.E., Jr. (1966b). Essay Examination Reliability. Unpublished MSS.

Harper, A.E., Jr. (1967) A Study of Examinations, Indian Educational Review, 2, 26-41.

Harper, A.E., Jr., Misra, V.S., & Gupta, R.L. (1967). Four Thousand Answer Books Remarked. Unpublished MSS.

Harper A.E. Jr. and Misra, V.S. (1976). Research On Examinations in India. NCERT, New Delhi.

Heywood, J. (1977). Assessment in Higher Education. John Wiley & Sons. London, New York, Sydney & Toronto.

Hill, Walker, H. (1963). How to Improve Essay Examinations. Teacher Education, 8, 8-9.

Hill, Walker, H. (1964a). Examinations for Secondary Education. Education and Psychology Review. Baroda. 4-8.

Hill, Walker, H. (1964b) Examination Improvement and Examination Reform. Journal of Education & Psychology, 22, 63-70.

Hill, Walker, H. (1965). What Research Says About Examinations. Teacher Education, 9, 1-10.

Hill, Walker H. (1967a). The Progress of Examination Reform (Part II). The progress of Education, 1-8.

Hill, Walker, H. (1967b). Why Examination System needs to be changed. Journal of Education & Psychology, 25, 20-27.

Hill, Walker, H. (1967c). The Role of the Training College in Examination Reform. Teacher Education, 1, 21-28.

Hoyt, C. (1935). Journal of Education Psychology.

Lele, P.T., Patel, P.M., Parikh, N.P., & Palkar, S.G.(1962). Qualitative Analysis of Essay Questions. Research Studies in Examinations. Baroda, December, 1962.

Mahalanobis, P.C. and Chakravarty, K.N. (1934). Analysis of Marks in the School Leaving Certificate Examination in the United Provinces, India. 1919. Sankhya 1. 245-266.

Mascarenhas, M. (1973). A Model Question Paper. Eduational India. 40 (3). September. p. 117-120.

Misra, V.S. (1968a). An Investigation into Admission Criteria. Gauhati. Examination Research Unit, Gauhati University.

Misra, V.S. (1968b). A Study of Essay Type Test Reliability. Gauhati. Examination Research Unit, Gauhati University.

Misra V.S. (1969). Study of the Effect of Randomisation and Scaling. On the Errors in Examination Marks. Gauhati. Examination Research Unit, Gauhati University.

Misra, V.S. (1976)- Given in Harper & Misra (1976).

Mudaliar Commission (1952). The Report of the Secondary Education Commission. New Delhi. Government.

Mukerjee, N. (1961). On Examinations. Journal of Gujarat Research Society (23).

Mukherjee, N.M. (1964). External Examinations. Calcutta: West Bengal Head Masters' Association. 338-343.

Natarajan, V. (1977) Monograph on Test & Item Analysis. AIU Publication.

S.S.C. Board, Maharashtra (1960). Statistical Investigations On Problems Pertaining to the Results of the S.S.C. Examination, Poona: S.S.C. Board.

Taylor, H.J. (1962a). Classifying Students by Examinations. Report on Examination Reform. UGC.

Taylor, H.J. (1962b). An Examination of Examiners: A Study of the Reliability of Examination Marking. Journal of the University of Gauhati.

Taylor, H.J. (1962c). Choice and Chance in Conventional Examinations. Report on Examination Reform. UGC.

Taylor, H.J. (1962d). Choice and Chance in Conventional Examinations. Report on Examination Reform. UGC., Appendix XII. (d).

Taylor, H.J. (1962e). A Hand Book of Scaling. Gauhati. Examination Research Unit, Gauhati University.

Taylor, H.J. (1963a). Operation Pass mark. Gauhati. Examination Research Unit, Gauhati University.

Taylor, H.J. (1963b). The Question Paper. General Education 4, 1-9.

Taylor, H. J. & Tluanga, L.N. (1963b). Scaling Tables. Gauhati. Examination Research Unit, Gauhati University.

Taylor, H.J., Tluanga, L.N. and Misra, V.S. (1966). A Study in Multiple Marking. Gauhati. Examination Research Unit, Gauhati University.

Tluanga, L.N., Taylor, H.J., Tiewsoh, F.N., Roy & Medhi, B.C. (1964). The Influence of Scaling on Examination Results. Gauhati: Examination Research Unit, Gauhati University.

University Education Commission (1950). Report of the University Education Commission. New Delhi: Government of India.

CHAPTER - 4

American Psychological Association (1974). Minutes of the Annual Meeting of the Counsul of Representatives.

Anastasi, A. (1976). Psychological Testing Macmillan Publishing Co., INC. New York, Collier Macmillan, London.

Bracht, G.H. & Hopkins, K.D. (1968). Objective and Essay Tests: Do they Measure different Abilities? Paper Presented at a Joint Session of the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

Coffman, W. (1966). On the Validity of Essay Tests of Achievement. Journal of Educational Measurement, 3, 151-156.

Dick, W. and Hagerty, N. (1971). Topics in Measurement-Reliability and Validity, McGraw-Hill, New York.

Ebel, R.L. (1979). Measuring Educational Achievement, Prentice Hall, U.S.A.

Gronlund, N.E. (1976). Measurement and Evaluation in Teaching (3rd ed.). Mac-millan, New York.

Harper, A.E., Jr. (1963). Objective and Traditional Examinations: Some Research. Education and Psychology Review, Spring, 63-78.

Holloway, P.J., Hardwick, J.L., Morris, J. and Start,K.B. (1967). The Validity of Essay and Viva-Voce Examining Techniques. British Dental Journal, 123 (8), 227-232. No.3, 141.

Hudgon, L. (1960). Degree Class and Attainment in Scientific Research. British Journal of Educational Psyghology; 51,67-73.

Kamat, A.R., & Deshmukh, A.C. (1961). Prediction Value of Marks at the Preceeding Examination in Forecasting Marks at the Succeeding Examination. Artha Vijnan.

Lele, P.T., Patel, P.M., Parikh, N.P., & Palkar, S.G.(1963). Empirical Studies in Examinations: Research Monograph-2, Baroda: Examination Reform and Research Unit, M.S. University, Baroda, March, 1963.

Lindzey, G. and Aronson, E. (1975).(Editors). A Hand Book of Social Psychology, Second Edition, Addison-Wesley Publishing Company, Inc.

Misra, G.S. (1964). Prognostic Value of High School and Intermediate Marks. Unpublished M.Ed. Dissertation, Allahabad University.

Patel Y.P. (1962). Statistical Analysis of Marks Obtained at the B.Ed. Examination (1961) of S.V. Vidyapeeth. Studies in Education.

Pidgeon, D. and Yates A. (1969). An Introduction to Educational Measurement, Routledge and Kegan Paul, London.

Popham, W.J.(1975). Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Taylor,HJ. (1962) The Marks of Examiners. Report on Examination Reform. UGC

Taylor, H.J. Tluanga, L.N., & Misra, V.S. (1966). Examination as Preditor. Gauhati: Examination Research Unit, Gauhati University.

Wiseman, S. (1956). The Use of Essays in Selection at 11+. III Reliability and Validity. Brisish Journal of Educational Psyyhology.

CHAPTER - 5

Bayley D.H. (1972). The Emptiness of Curriculum Reform. Journal of Higher Education, 44, (8), 591.

BITS Bulletin (1980-81). Published by BITS, Pilani.

Carnegie Commission on Higher Education (1973). The Purposes and Performance in Higher Education in the United States. Approaching the Year 2000, McGraw-Hill, New York.

Furst, E.J. (1958). Constructing Evaluation Instruments, David Mackay, New York.

Heywood, J. (1977). Assessment in Higher Education. John Wiley & Sons. London, New York, Sydney, Toronto.

Iyer T.S.K.V. and Krishnamurthy V. (1976). Examination Reform-Innovations in Teaching and Learning. Journal of Higher Education, Vol.2 No.1. University Grants Commission, New Delhi.

Jevons, F.R. (1972). In F.R. Jevons and H.D. Turner (Eds.) What kind of Graduates do we need? Oxford U.P., London.

Mandke, V.V. (1975). Practice School - Linking Higher Education with Environment. CSIO Communications. Vol. No.4, Page 62-67.

Mandke, V.V. (1976). Engineering Education in India: Some Directions and Some Experiments. Paper Presented at IIT,Madras on the Occasion of the Celebration of 25th Anniversary of IIT Kharagpur (Feb. 1976).

Mandke, V.V. (1978). A Note on the Role of a University in Serarch of Techniques for the Rural Poor of Third World. A Paper presented at International Seminar on 'Techniques for the Rural Poor of the Third World' held at Centre of Science for Villages at Sevagram, Wardha from 22 to 25 Jan, 1978.

Mandke, V.V. (1980). BITS Practice School - A Case Study in Industry University Collaboration, BITS, Pilani Publication.

Mitra, C.R. (1974). Some New Directions for Higher Education. Swarajya Weekly. Madras, May.

Mitra, C.R. (1975). Education for Self-Reliance in India. Presented at the Seminar-Cum-Workshop on Alternative Technology held at Simla in September, 1975 under the Joint auspices of C.S.I.R. and the Indian Institute of Advanced Study, Simla.

Sherrin, D.C. and T.R. Long (1972). The Educator's Dilemma: What Makes Clyde Want to Learn. J. Engng. Ed., 63 (3), 188.

Whitehead, A.N. (1933). The Aims of Education. Benn, London.

CHAPTER - 6

Academic Regulations (1977). BITS Publication

Anastasi, A. (1976). Psychological Testing, Macmillan Publishing Co., Inc. New York, Collier Macmillan, London.

Guidelines for PS-I Operation and Evaluation (1979). BITS Publication.

Gullicksen, H. (1950). Theory of Mental Tests. New York. John Wiley and Sons.

Mandke, V.V. (1980). BITS Practice School - A Case Study in Industry University Collaboration. BITS, Pilani Publication

CHAPTER - 7

BITS Bulletin (1980-81). BITS Publication.

Heiss, A. (1973). An Inventory of Academic Innovation and Reform, Canegi Commission on Higher Education, Barkeley.

Heywood, J. (1977). Assessment in Higher Education. John Wiley & Sons. London, New York, Sydney, Toronto.

Livingstone, J.S. (1971). The Myth of the Well-Educated Manager. Harvard Business Review, Jan/Feb.

M.E.(Collaborative) Planning - Office Files on M.E.(Collaborative) Programme, PS Division, BITS, Pilani.

Saupe, J. (1961). In P. Dressel (Ed.) Evaluation in Higher Education, Houghton Mifflin, Boston.

Whitehead, A.N. (1933). The Aims of Education. Benn, London.

Appendix A

June'80

Questionnaire I for finding out the Objectives of Education:-

Name of the teacher______________________________

Courses taken in the recent past ______________________________

____

____

____

Following list contains some of the qualities which are 'desired to be common among most adult situations. They are believed to be essential for living. How would you rate these qualities ? Indicate your opinion as to which of the qualities should be achieved by university education. What qualities can be achieved by education system like BITS and finally which of the qualities you have been able to achieve in your course/courses ?

Sd/-
(ANITA MISRA)
Research Student
R & C Division
BITS, Pilani

S. No.	Desired Qualities	Desirability Rating	Qualities that should be achieved by education.	Qualities that can be achieved	Qualities that you have been able to achieve	REMARKS
1.	Perception					
2.	Analysis					
3.	Diagnosis					
4.	Problem solving					
5.	Judgement					
6.	Communication					
7.	Understanding					
8.	Sympathy					
9.	Tolerance					
10.	Sense of responsibility					
11.	Leadership					
12.	Decision making					
13.	Coping with frustration					
14.	Knowledge					
15.	Self reliance					
16.	Creativeness					
17.	Confidence in one's own abilities.					
18.	Sense of Humour					
19.	Ability to mix well					
20.	Wide interest					

S. No.	Desired Qualities	Desirability Rating	Qualities that should be achieved by education	Qualities that can be achieved	Qualities that you have been able to achieve	REMARKS
21.	Independence					
22.	Honesty					
23.	Ambition					
24.	Common sense					
25.	Logical thought					
26.	Any other (specify)					

Appendix B

Questionnaire I for finding out the Objectives, Techniques and Nature of Examinations at BITS

June'80

Respected Sir/Madam,

I am conducting this survey to find out about the objectives, techniques and nature of evaluation in the examinations at BITS for my doctoral work. I shall be grateful if you could please fillup the enclosed questionnaire and send it back to me as soon as convenient. In filling up the questionnaire please refer to a particular course which you have taken in recent past. Please give the name of the course at the appropriate place on this page.

With regards,

Yours sincerely,
Sd/-
(ANITA MISRA)

Please return the completed questionnaire including this page to -

Miss ANITA MISRA
Research Student
R & C Division
BITS, Pilani-333031

This questionnaire has three parts:

Part A	Objectives of evaluation
Part B	Techniques of evaluation
Part C	Nature of examinations

Your name please ______________________

Title of the course ______________________

Name of the Instructor Incharge ______________________

Course Number ______________________

Semester/Year ______________________

Units ______________________

Approximate number of students in the whole course ______________________

Part A : Objectives of Evaluation

Q. 1 We have listed below some of the objectives of evaluation. In your opinion, which of the objectives are being achieved with respect to the course you have conducted ? Please indicate the weightage assigned to these objectives.

(For explannatory notes see below)

S.No.	Objective	Quizzes	Tests	Compre-hensive	Others*
I	Knowledge				
II	Comprehension				
III	Application				
IV	Analysis				
V	Synthesis				
VI	Evaluation				

Explamnatory notes for Question 1:

*Others - This includes components like home assignment, seminar, viva, group discussion, project report etc.

I Knowledge

Knowledge includes those test situations which emphasize the remembering either by recognition or recall. In the learning situation the student is expected to store in his mind certain information, and behaviour expected later is the remembering of this information. Knowledge includes several sub-categories such as knowledge of terminology, knowledge of specific facts (dates, events, persons, places etc.), knowledge of classification and categories, knowledge of methodology (methods of inquiry, techniques and procedures), knowledge of principles, theories and so on.

II Comprehension

Comprehension refers to the type of understanding so that the individual knows what is being communicated and can make use of the communication or idea without necessarily relating it to the other material or seeing its fullest implications. Categories included in the comprehension are :-

(a) Translation :- It is the ability to change the communication in his mind in parallel form, e.g. ability to translate mathematical verbal statements into symbolic form and vice-versa.

(b) Interpretation:- Whereas translation involves an objective part for rendering or communication, interpretation involves a reordering, rearrangement or a new view of the material.

(c) Extrapolation :- The ability to make simple extensions beyond what is given in the communication itself is called extrapolation, e.g. extension of some trends or tendencies beyond the given data.

III Application :-

If testing situations are to involve application, they must either be situations new to student or situations containing new elements as compared to the situation in which particular abstract was learnt. If the test situations are similar in which abstraction was learnt, then it would not be called 'application'. It would simply be 'recall' or 'knowledge'. Ideally in 'application' we should seek a problem which will test the extent to which individual has learnt to apply the abstraction, e.g. application of technical principles, ideas and theories which must be remembered and applied.

IV Analysis

Analysis includes breakdown of a communication into its constituent elements or parts and detection of the relationships of the parts and the way they are organized. Some of the important classes of the analysis are:-

(a) Analysis of elements in communication, e.g. analysis of data.

(b) Analysis of relationships e.g. ability to understand the connections and interactions between elements and parts of a communication.

V Synthesis

Synthesis includes putting together elements and parts so as to form a whole. In other words it includes planning of some idea. In synthesis student must draw upon elements from many sources and put them together into a structure or pattern not clearly seen there before, e.g. production of a unique communication/writing.

VI Evaluation

Evaluation involves making judgements either qualitative or quantitative e.g. ability to make comparisons.

Q. 2 Apart from the above objectives, do you think some other qualities like leadership, sense of responsibility, decision-making, co-operation etc. can be measured by class-room evaluation ?

Yes	No

Q. 3 If your answer to the above question is yes, what techniques of evaluation would you include in evaluation in order to achieve them ?

____

____

____

---- 0 ---- 0 ---- 0 ----

Part B - Techniques of evaluation

Q. 1 How many quizzes and tests do you conduct in the course in a semester ?

Component	No. of Components	Weightage
Quizzes		
Tests		
Comprehensive		

Q.2 Apart from quizzes and tests what other instruments of evaluation do you incorporate in the evaluation ? (Please indicate weightage)

Home Assign- ment	Lab work	Project work	Viva	Seminar	Group dis- cussion	Any other	None

Q. 3 Please identify and explain the objectives of the evaluation method referred to in Q.2.

____

____

____

____

____

---- 0 ---- 0 ---- 0 ---- 0 ----

Part C - Nature of Examinations

Q.1 What type of examinations do you give as far as 'open book' and 'closed book' examinations are concerned ?

	Components		
Type of examina- tion	Quizzes	Tests	Compre- hensive
All/mostly Open Book			
All partly Open/Closed Book			
All/mostly Closed Book			

Q. 2 Do you give choice of questions or marks in the examinations ?

Components / Choice	Quizzes	Tests	Comprehensive
Include choice			
Do not include choice			

Q. 3 What is your opinion about giving choice in examinations ?

Components / Options	Quizzes	Tests	Comprehensive
Should include choice			
Should not include choice			
May or May not include choice			

Q. 4 What type of questions do you include in the question papers ? Please indicate weightage against each type of questions.

Components / Type of questions	Quizzes	Tests	Comprehensive
Objective type			
Short answer descriptive			
Short answer problem solving			
Long answer descriptive			
Long answer problem solving			
Any other (Please specify)_________			
____________________			
____________________			

APPENDIX 'C'

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI (RAJASTHAN)
EVALUATION SCHEME FOR PRACTIC SCHOOL -II

Practice Station : __________
Name of the Student : __________
ID Number : __________

Academic Year : __________
Student Discipline : __________
Practice Component Attended : __________
Practice Course No. __________

EVALUATION METHOD / STUDENT CHARACTERISTICS	QUIZZ		VIVA			SEMINAR				GROUP DISCUSSIONS				PROJECT REPORT				OBSERVATION	DIARY		TOTAL
	Knowledge	Application	Knowledge	Originality	Expression	Technical Credibility	Expression	Communication	Organisation of material	Intellectual Ability	Creativity	Approach to Problem Solving	Group Behaviour	Technical Credibility	Methodology	Documentation	Recommendations				
1. Knowledge of Concepts 1.1 With which he is already familiar	2		1			3								3						9	
1.2 Newly introduced at Practice School	4																			4	
1.3 Depth of Knowledge			1											1						2	18
2. Application of Principles 2.1 In a given situation		5				2														7	7
3. Intellectual Ability 3.1 To comprehend & act in new situations		1				1				2										4	
3.2 To follow logical path in problem solving		1	1												3					5	9
4. Creativeness & Originality 4.1 Conceiving new and unusual ideas	1			1							3				3					8	
4.2 Suggesting Practical and Good Solutions											1						3			3	9
5. Professional Judgement & Decision Making Ability 5.1 In evaluating alternatives.						2											3			5	5
6. Interdisciplinary Approch 6.1 Broad-based Knowledge						1									1			1		3	
6.2 In problem Solving situation															2			1		3	6
7. Skills for Data Handling 7.1 Understanding of data gathering/processing techniques												1			1					2	
7.2 Choice of Measurement tools												1			1					2	
7.3 Skills for hardware implementation																		1		1	5
8. Documentation 8.1 Review of literature						1										2				3	
8.2 Organising the material									1							2			1	4	7

EVALUATION METHOD / STUDENT CHARACTERISTICS	QUIZE		VIVA			SEMINAR				GROUP DISCUSSIONS				PROJECT REPORT				OBSERVATION	DIARY		TOTAL
	Knowledge	Application	Knowledge	Originality	Expression	Technical Credibility	Expression	Communication	Organisation of material	Intellectual Ability	Creativity	Approach to Problem Solving	Group Behaviour	Technical Credibility	Methodology	Documentation	Recommendations				
9. Self-Expression 9.1 Conveying ideas clearly					1															1	
9.2 Delivery & style of presentation							2													2	
9.3 Language							2									2				4	
9.4 Black-board presentation								1												1	
9.5 Platform manners								1												1	
9.6 Introducing & ending of the presentation								1												1	10
10. Initiative 10.1 In taking lead in problem solving situations.													2					1		3	3
11. Self-Reliance 11.1 Confidence in one's own abilities																		3		3	3
12. Co-operation : 12.1 With group/Instructor & organisation													2					1		3	3
13. Leadership 13.1 Organizing the efforts of the group						1												2		3	
13.2 Moderating discussions								1					2							3	
13.3 Inspiring the group members																		1		1	7
14. Industry : 14.1 Making forceful efforts to know more																		1		1	
14.2 Desire to exceed minimum effort expectation																			2	2	3
15. Sense of Responsibility 15.1 Reliability in meeting problem objectives															1					1	
15.2 Planning & meeting deadlines																		1	1	2	
15.3 Punctuality																		1	1	2	5
16 Social Sense 16.1 Ability to create good impression and to act accordingly						1							2					1		3	3
TOTAL	8	7	2	1	1	11	4	4	1	2	3	2	8	3	11	6	8	15	5		100
	15		5			20				15				28				15	5		

Date : ____________

Signature of Instructor-in-charge

APPENDIX D

A Typical PS-I Evaluation Scheme as Implemented at Khetri in Summer, 1973.

Evaluation Scheme	Marks
1. Assignments given in terms of 'Exercises' were evaluated in the following manner:	
(a) Report Writing (i) Exercises 1 to 4 of the orientation type spread over total duration of 20 days. Evaluation on group basis.	20
(ii) Exercises from 5 to 10. Here ideas are of importance. Total duration 15 days. Evaluated on individual basis.	15
(b) Diary. Every student was required to keep a diary which consisted of day to day observations.	05
2. Oral presentation of Group Reports. The exercise work tackled in groups. Each group had a leader who had total responsibility of the assignment. It is the leader who also had the responsibility of presenting the 'Exercise'. During the entire summer term one student gets about four chances to present oral reports.	10
3. Participation. Every student's participation in his own group's work as well as in the activities of other groups was judged.	20
4. Personal Characteristics. During the discussions which could be arranged for an hour almost daily every student was evaluated for the following characteristics. (a) Leadership (b) Sense of reponsibility (c) Cooperation (d) Initiative (e) Regularity and progress of the group	15
5. Quizzes and Short Reports. In two months about 2 quizzes and 7 short reports were written based on visits to various departments of the organisation.	15
TOTAL MARKS	100

APPENDIX E

A Typical PS-I Evaluation Scheme as Implemented at Khetri during summer 1974.

Evaluation Scheme	Marks
1. Quiz on 'Know your Organisation'	10
2. Quiz on Gap Lectures	10
3. Assignment Evaluation (Except for the last assignment)	
(i) Knowledge and application of scientific fundamentals	12
(ii) Knowledge of technological operations	12
(iii) Oral Presentation	10
(a) Self-expression	
(b) Material organisation	
(c) Black-board presentation	
(d) Participation	
(e) Platform manner	
(iv) Written Presentation	12
(a) Presentation scheme	
(b) Preciseness	
(c) Logical development of argument	
(d) Force of expression	
(v) Sense of responsibility	03
(vi) Initiative	03
(vii) Co-operation	03
(viii) Leadership quality	03
(ix) Industry	02
4. Open-ended project, i.e. Final assignment evaluated for 'Ideas'	10
5. Diary. Student is required to keep a diary which incorporates his day to day observations	10
Total Marks	100

APPENDIX F

Practice-II Evaluation Scheme as Implemented during 1973-75

I. Weightage

(i)	Know your factory (Orientation)	5%
(ii)	Gap Lectures	5%
(iii)	Projects (to be equally weighted)	90%

II. Evaluation Scheme

(a) Know your factory and gap lectures will be evaluated through quiz/viva.

(b) Each project will be evaluated once/twice during the course of the project and at the end of the project in respect of the following items with weightages indicated against each.

1.	(i)	Knowledge and application of engineering fundamentals	10
	(ii)	Engineering judgement and decision making	10
	(iii)	Leadership	04
	(iv)	Co-operation	04
	(v)	Sense of responsibility	04
	(vi)	Initiative	04
	(vii)	Industry	04
2.		Quiz/viva conducted during the course of the project	10
3.		Oral Presentation* of the Progress Report during the course of the project.	10
4.	(i)	Written Presentation**	20
	(ii)	Conclusions and recommendations	10
	(iii)	Final oral presentation	10
		Total	100

---

* Oral Presentation will be judged in respect of:

(i) Self-expression, (ii) Material organisation, (iii) Black-board organisation, (iv) Technical quality of answers, (v) Participation, (vi) Platform manners.

** Written presentation will be judged in respect of:

(i) Presentation scheme, (ii) Preciseness

(iii) Logical development of argument, (iv) Force of expression.

APPENDIX G

RATING SHEET

PRACTICE COURSE ________

(To be filled in triplicate. One copy to be retained at Practice Station and remaining two to be returned to Practice School Division Office along with student grade)

Name ______________ ID No. ________ Year _____ Semester ______

Practice Station __________ Degree _________ Discipline ______

Please rate the student by tick mark ( ) in the appropriate column.

Personality Traits	Excellent	Good	Average	Poor	V.Poor
1. Knowledge and Application of Fundamental Principles					
2. Intellectual Ability					
3. Creativity and Art of Guestimation					
4. Professional Judgement					
5. Problem Solving Ability					
6. Decision-making Ability					
7. Ability to Communicate					
8. Initiative and Self-reliance					
9. Team work					
10. Leadership					
11. Punctuality and Ability to Meet Deadlines					
12. Sense of Responsibility and Common Sense					

Please check traits which best describe personality

Confident Poised & Curteous	Should be less aggressive	Pleasant curious & forceful	Should be friendlier with group	Likeable	Should be more aggressive

Indicate work for which he is best suited. Check only one or indicate order of choice.

Research ______ Development ______ Teaching _____ Design _____

Production _____ Sales & Marketing _____ Finance _____ EDP ______

Scientist _____ Journalist _______ Other __________________

If necessary, you may elaborate on your reasons for above ratings and add any further comments you may have. Use the back of this sheet, if needed.

Grade Obtained at Practice Course ________

Signature of Instructor-in-charge

Date __________

Name ______________

APPENDIX H

Academic Regulations – **CLAUSE 22.2**

In all practice school courses also the continuous evaluation enunciated in clause 13 of Academic Regulations will be followed. Since the educational processes in the practice school courses seek out and focus attention on many latent attributes which do not surface in the normal classroom situation, the process of evaluation in the practice school courses should be designed with care so that information on a continuous basis on the following attributes becomes available: intellectual ability; team work; leadership; initiative; personality; professional judgement; common sense; problem solving ability; sense of responsibility; decision making ability; art of guestimation; punctuality; ability to meet deadlines; ability to communicate through oral and written presentations, etc. Each such student may also be given a statement describing qualitatively the degree to which these attributes have been demonstrated by him in the course.

APPENDIX I

(i) Program for Correlation Coefficients

```
// JOB T                                                  ANITA
// FOR
*LIST SOURCE PROGRAM
*IOCS(2501 READER,1403 PRINTER)
*ONE WORD INTEGERS
C
C
C-----PROGRAM FOR CORRELATION COEFFICIENTS
C-----THIS PROGRAM ALSO GIVES  F  AND  T VALUES
C-----CORRELATION NUMBERS PRINTED IN MATRIX FORM BETWEEN TESTS
C
C
      REALMQ(10)
      COMMON XM(400,10),C(10,10),T(10,10),P(10,10),PM(10),SM(10),CM(10)
C
C
C-----READ DATA
C-----NC = NUMBER OF STUDENTS
C-----NT = NUMBER OF TESTS
C-----PM(J) = MAXIMUM MARKS IN J TH QUESTION
C-----XM(I,J) = MARKS SCORED BY I TH STUDENT IN  J TH  TEST
C
C
      READ(8,20)NC,NT
      WRITE(5,20) NC,NT
      CN=NC
      ACN=1./(CN*(CN-1.))
      TN=NT
      READ(8,21)(PM(J),J=1,NT)
      WRITE(5,21)(PM(J),J=1,NT)
      DO 1 J=1,NT
      SM(J)=0
      PM(J)=100./PM(J)
      CM(J)=0
    1 CONTINUE
      WRITE(5,23)
      DO 3I=1,NC
      READ(8,50) M,(MQ(J),J=1,NT)
      WRITE(5,22)I,(MQ(J),J=1,NT)
      DO 2 J=1,NT
      XM(I,J)=MQ(J)
      XM(I,J)=XM(I,J)*PM(J)
      SM(J)=SM(J)+XM(I,J)
      CM(J)=CM(J)+XM(I,J)**2
    2 CONTINUE
    3 CONTINUE
      GM=0
      DO 4 J=1,NT
      SM(J)=SM(J)/CN
      GM=GM+SM(J)
    4 CONTINUE
      GM=GM/TN
```

```
      WRITE(5,24)
      DO 5 I=1,NC
      WRITE(5,22)I,(XM(I,J),J=1,NT)
      DO 5 J=1,NT
      XM(I,J)=XM(I,J)-SM(J)
    5 CONTINUE
      WRITE(5,21)(SM(J),J=1,NT)
      WRITE(5,30)GM
      WRITE(5,21)(CM(J),J=1,NT)
      DO 6 M=1,NT
      DO 6 N=1,NT
      T(M,N)=0
    6 P(M,N)=0
      DO 7 I=1,NC
      DO 7 M=1,NT
      DO 7 N=M,NT
      P(M,N)=XM(I,M)*XM(I,N)+P(M,N)
    7 CONTINUE
      DO 8 M=1,NT
      DO 8 N=M,NT
      A=P(M,M)*P(N,N)
      A=SQRT(A)
      C(M,N)=P(M,N)/A
      C(N,M)=C(M,N)
    8 CONTINUE
C     CALCULATION OF F
      VB=0
      VW=0
      DO 9 M=1,NT
      VW=VW+P(M,M)
      S=SM(M)-GM
      VB=S*S+VB
    9 CONTINUE
      VW= VW /(TN*CN-TN)
      VB=CN*VB/(TN-1.)
      F=VB/VW
      WRITE(5,25)F
C     CALCULATION OF T NUMBER
      DO 11 M=1,NT
      DO 11 N=M,NT
      A=1.-C(M,N)*C(M,N)
      IF(A)90,91,90
   90 CONTINUE
      A=1./A
      A=A*(CN-2.)
      A=SQRT(A)
   91 CONTINUE
      T(M,N)=A*C(M,N)
      T(N,M)=T(M,N)
   11 CONTINUE
      WRITE(5,26)
      DO 12 M=1,NT
```

```
      WRITE(5,28)(C(M,N),N=1,NT)
   12 CONTINUE
      WRITE(5,27)
      DO13M=1,NT
      WRITE(5,28)(T(M,N),N=1,NT)
   13 CONTINUE
      WRITE(5,29)
   20 FORMAT(2I5)
   21 FORMAT(4F10.2)
   22 FORMAT(I5,4F10.2)
   50 FORMAT(I5,4F6.2)
   23 FORMAT(15H ORIGINAL MARKS)
   24 FORMAT(17H PERCENTAGE MARKS)
   25 FORMAT(11H F NUMBER  ,F20.4)
   26 FORMAT(17H CORR COEFFICIENT)
   27 FORMAT(14H T TEST VALUES)
   28 FORMAT(3X///10E11.2)
   29 FORMAT(31H FINAL RESULT IS THREE BY EIGHT)
   30 FORMAT(12H GRAND MEAN ,F20.2)
      DO40I=1,NT
      DO40J=I,NT
      YZ=(P(I,I)+P(J,J))*ACN
      YZ=SQRT(YZ)
      T(I,J)=(SM(I)-SM(J))/YZ
      T(J,I)=T(I,J)
   40 CONTINUE
      DO41I=1,NT
   41 WRITE(5,28)(T(I,J),J=1,NT)
      CALL EXIT
      END
// XEQ
```

(ii) Analysis of Variance Program for Test Reliability and Course Reliability

```
// JOB T                                                        ANITA
// FOR
*LIST SOURCE PROGRAM
*ONE WORD INTEGERS
*IOCS(2501READER,1403PRINTER)
C
C
C-----ANALYSIS OF VARIANCE PROGRAM FOR TEST RELIABILITY AND COURSE
C-----RELIABILITY
C-----REFERENCE THREE BY EIGHT
C-----OUTPUT GIVES VARIOUS SUM OF SQUARES AND RELIABILITY
C-----RTT DENOTES RELIABILITY
C-----VE AND VR DENOTE TOTAL AND ERROR VARIANCE
C
C
      DIMENSION R(400),C(35),QM(35)
C
C
C-----READ INITIAL DATA
C-----NS = NUMBER OF STUDENTS
C-----NQ = NUMBER OF QUESTIONS
C
C
      READ(8,1)NS,NQ
      WRITE(5,11)NS,NQ
      SN=NS
      QN=NQ
      SS=1./(SN-1.)
      QQ=1./(QN-1.)
      SN=1./SN
      QN=1./QN
      SQN=SN*QN
      SQ=SS*QQ
      X=0
      XR=0
      XC=0
      XRR=0
      XCC=0
    1 FORMAT(2I5)
      DO2I=1,NS
      R(I)=0
    2 CONTINUE
      DO 3J=1,NQ
      C(J)=0
    3 CONTINUE
      DO 5I=1,NS
C
C
C-----READ DATA ON MARKS
C-----QM(J) = MARKS SCORED IN J TH QUESTION
C
C
```

```
      READ(8,4) (QM(J),J=1,NQ)
      WRITE(5,17)I,(QM(J),J=1,NQ)
4     FORMAT(5X,7F6.2)
      DO 5 J=1,NQ
      Y=QM(J)
      R(I)=R(I)+Y
      C(J)=C(J)+Y
      X=X+Y*Y
    5 CONTINUE
      WRITE(5,12)
      WRITE(5,13)(C(J),J=1,NQ)
      WRITE(5,14)
      DO 15I=1,NS
      WRITE(5,16)I,R(I)
   15 CONTINUE
      DO 6 I=1,NS
      Y=R(I)
      XRR=XRR+Y*Y
    6 XR=XR+Y
      DO 7 I=1,NQ
      Y=C(I)
      XCC=XCC+Y*Y
    7 XC=XC+Y
      WRITE(5,8)XR,XC
    8 FORMAT(27H ROW SUM AND COLUMN SUM    =,2F20.4)
      XRC=XR*XR*SQN
      SDE=XRR*QN-XRC
      SDI=XCC*SN-XRC
      TSS=X-XRC
      RSS=TSS-SDE-SDI
      VE=SDE*SS
      VR=RSS*SQ
      RTT=1.0-VR/VE
      WRITE(5,9)SDE,SDI,TSS,RSS
    9 FORMAT(19H SDE SDI TSS RSS    ,4F15.6)
      WRITE(5,10)VE,VR,RTT
   10 FORMAT(13H VE VR RTT    ,3F20.6)
   11 FORMAT(36H NUMBER OF STUDENTS AND QUESTIONS    ,2I10)
   12 FORMAT(12H COLUMN SUMS)
   13 FORMAT(5X,4(10F10.3/))
   14 FORMAT(9H ROW SUMS)
   16 FORMAT(I10,F20.4)
   17 FORMAT(I5,4(10F10.3/))
      CALL EXIT
      END
// XEQ
```

(iii) Analysis of Variance Program for Calculating Reliability of Examinations where choice of questions is allowed.


```
// JOB T                                                    ANITA
// FOR
*LIST SOURCE PROGRAM
*IOCS(2501READER,1403PRINTER)
C
C
C-----ANALYSIS OF VARIANCE PROGRAM FOR CALCULATING
C-----RELIABILITY OF EXAMINATIONS, WHERE CHOICE
C-----OF QUESTIONS IS ALLOWED
C
C
      DIMENSION R(400),C(35),QM(35)
      READ(8,1)NS,NQ,NQT
      WRITE(5,11)NS,NQ,NQT
      SN=NS
      QN=NQ
      SS=1./(SN-1.)
      QQ=1./(QN-1.)
      SN=1./SN
      QN=1./QN
      SQN=SN*QN
      SQ=SS*QQ
      X=0
      XR=0
      XC=0
      XRR=0
      XCC=0
1     FORMAT(3I5)
      DO2I=1,NS
      R(I)=0
    2 CONTINUE
      NQ=NQT
      DO 3J=1,NQ
      C(J)=0
    3 CONTINUE
      DO 5I=1,NS
      READ(8,4) (QM(J),J=1,NQ)
      WRITE(5,17)I,(QM(J),J=1,NQ)
    4 FORMAT(5X,7F5.1)
      DO 5 J=1,NQ
      Y=QM(J)
      R(I)=R(I)+Y
      C(J)=C(J)+Y
      X=X+Y*Y
    5 CONTINUE
      WRITE(5,12)
      WRITE(5,13)(C(J),J=1,NQ)
      WRITE(5,14)
      DO 15 I=1,NS
      WRITE(5,16)I,R(I)
   15 CONTINUE
      DO 6 I=1,NS
```

```
      Y=R(I)
      XRR=XRR+Y*Y
    6 XR=XR+Y
      DO 7 I=1,NQ
      Y=C(I)
      XCC=XCC+Y*Y
    7 XC=XC+Y
      WRITE(5,8)XR,XC
    8 FORMAT(27H ROW SUM AND COLUMN SUM    =,2F20.4)
      XRC=XR*XR*SQN
      SDE=XRR*QN-XRC
      SDI=0
      TSS=X-XRC
      RSS=TSS-SDE-SDI
      VE=SDE*SS
      VR=RSS*SQ
      RTT=1.0-VR/VE
      WRITE(5,9)SDE,SDI,TSS,RSS
    9 FORMAT(19H SDE SDI TSS RSS    ,4F15.6)
      WRITE(5,10)VE,VR,RTT
   10 FORMAT(13H VE VR RTT    ,3F20.6)
   11 FORMAT(36H NUMBER OF STUDENTS AND QUESTIONS    ,3I10)
   12 FORMAT(12H COLUMN SUMS)
   13 FORMAT(5X,4(10F10.3/))
   14 FORMAT(9H ROW SUMS)
   15 FORMAT(I10,F20.4)
   17 FORMAT(I5,4(10F10.3/))
      CALL EXIT
      END
// XEQ
```

(iv) Program for Analysis of PS-II Evaluation Based on Students' Feed-back.


```
// JOB T                                                    ANITA
// FOR
*LIST SOURCE PROGRAM
*IOCS(2501READER,1403PRINTER)
C
C
C-----PROGRAM FOR ANALYSIS OF PS II EVALUATION
C-----BASED ON STUDENTS' FEED BACK
C
C
      DIMENSION JM(36,138),MJ(7,138),X(7).
      NI=36
      NQ=138
      MQ=7
      FM=100./36.
      DO1 I=1,NQ
      DO1J=1,MQ
      MJ(J,I)=0
    1 CONTINUE
      DO5I=1,NI
      READ(8,2)IS,IN,ID,(JM(I,J),J=1,64)
      READ(8,3)(JM(I,J),J=65,NQ)
      WRITE(5,4)IS,IN,ID,(JM(I,J),J=1,NQ)
    2 FORMAT(2I2,32I1,I2,5I3,27I1)
    3 FORMAT(5X,74I1)
    4 FORMAT(3I3,31I3,/,I50,5I5,/,3(I12,34I3/))
    5 CONTINUE
      DO7J=1,NI
      NA=1
      NB=31
      DO7IR=1,2
      DO6IN=NA,NB
      I=JM(J,IN)+1
      MJ(I,IN)=MJ(I,IN)+1
    6 CONTINUE
      NA=38
      NB=NQ
    7 CONTINUE
      DO9I=1,NQ
      DO8J=1,MQ
      Y=MJ(J,I)
    8 X(J)=Y*FM
      WRITE(5,10)I,(MJ(J,I),J=1,MQ)
      WRITE(5,11)(X(J),J=1,MQ)
   10 FORMAT(I5,7I10)
   11 FORMAT(F15.2,6F10.2/)
    9 CONTINUE
      CALL EXIT
      END
// XEQ
```