CHAPTER 2
REVIEW OF RELATED LITERATURE
CHAPTER 2

REVIEW OF RELATED LITERATURE

2.1.0 OVERVIEW

The first section of this chapter presents the attributes associated with the concept of curriculum and the distinction between 'Intended curriculum' and 'Operational Curriculum'. The scope and concerns of curriculum studies are illustrated in the second section using a system's model and a paradigm of curriculum research and its categories. The third section portrays conceptual map of curriculum evaluation and presents a review of curriculum evaluation studies. The methodologies used for assessing the effectiveness of curricula are also reviewed. The use of criterion referenced tests for assessing the effectiveness of curricula are described in detail. Research using environment assessments in curriculum evaluation are also reviewed. The last section of this chapter deals with the research studies which formed the basis for selection of factors hypothesized to influence the effectiveness of chemistry curriculum of polytechnics. A summary of the directions provided by the literature review is given at the end of this chapter.

2.2.0 CONCEPT OF CURRICULUM

The word "Curriculum" comes from a Latin root meaning "race course" or the ground to be covered to reach a goal. Traditionally curriculum was regarded as the relatively
standardized ground covered by students in their race toward the finish line (a diploma).

The literature on curriculum is rich with various definitions. The Dictionary of Education (Good, 1945) defines curriculum as follows:

i) a systematic group of courses or sequences of subjects required for graduation or certification in a major field of study;

ii) a general overall plan of the content or specific materials of instruction that the school should offer the student by way of qualifying him for graduation or certification or for entrance into a professional or vocational field;

iii) a body of prescribed educative experiences under school supervision, designed to provide an individual with the best possible training and experience to fit him for the society of which he is a part or to qualify him for a trade or a profession.

The word "curriculum" is used with a variety of meanings. According to Ochs (1974) "the term curriculum is often used to designate equally a programme for a given subject matter and for a given grade, a programme for a given subject matter for the entire course (study cycle) or the whole programme of different subjects for the entire course". Further, the term 'curriculum' is sometimes used in a wider sense to cover the various educational activities through which the content is conveyed as well as materials used and methods employed.
An analysis of the various definitions of curriculum reveals some of the following important attributes associated with the concept of curriculum.

- Related to an occupation
- Objective-oriented content
- Planned learning experiences
- Criteria for evaluation of student's performance.

Zais (1976, p.6-11) has classified the various definitions of curriculum into six viewpoints listed below.

i) Curriculum as the programme of studies
ii) Curriculum as course content
iii) Curriculum as planned experiences
iv) Curriculum as experiences "had" under the auspices of the school
v) Curriculum as a structured series of intended learning outcomes
vi) Curriculum as a (written) plan for action.

Taba (1962, p.9) finds the extreme breadth of the 'experience' definitions of the curriculum nonfunctional; on the other hand she feels that "excluding from the definition of curriculum everything except the statement of objectives and content outlines and relegating anything that has to do with .... learning experiences to 'method' might be too confining to be adequate for a modern curriculum." Her response to this dilemma of curriculum definition lies somewhere in between these two extremes. The suggestion and the central thrust of Taba's conception of curriculum is that the broader (i.e., more general)
aspects of purposes, content, and method belong in the realm of curriculum, while the more proximate and specific aspects properly are allocated to teaching and instruction. Taba's conception of curriculum does not employ an "implementation" criterion; rather, it depends on a relatively flexible and subjective judgement as to where a dividing line is to be drawn on a continuum which is clearly ultimate - general at the "Curriculum" pole and immediate - specific at the "Instruction" end. (See Figure 2.1)

**Figure 2.1**

Continuum along which subjective judgments are made to determine the curricular or instructional nature of educational phenomena

Ultimate-general  Immediate-specific

CURRICULUM  INSTRUCTION

Source: Zais, (1976, p.12.)

This relatively loose criterion is useful because it can be employed whether we think of curriculum as a document or as a cluster of phenomena in a live classroom situation. Summing up we can say that the curriculum provides direction for classroom
instruction, but it does not consist of a series of lesson plans. It is the teacher's prerogative and responsibility to interpret and translate the curriculum document in terms of his own and his students' experience.

The Curriculum Plan is implemented through the medium of instruction. Taylor and Colin (1979) have distinguished between 'intended curriculum' which refers to the prescriptions in the curriculum document and "the curriculum in operation or operational curriculum". When an "intended curriculum" is enacted in a classroom or given life through teaching it becomes an 'operational curriculum'. The 'intended curriculum' is an inert document containing the objectives of the curriculum, content matter, timeschedules, guidelines to teaching and learning and the performance standards expected. The "operational curriculum" deals with the processes of teaching and learning, organization of the class and the milieu in which instruction takes place.

Many factors influence/constrain teacher's efforts to implement the intended curriculum and make it a functional or operational curriculum. This may result in certain gaps between intention and realization.

The proposed research study aims to assess the instructional environment in which the curriculum is implemented in order to identify the factors affecting the effective implementation of the curriculum.
2.3.0 SCOPE AND CONCERNS OF CURRICULUM RESEARCH

Curriculum as a field of systematic inquiry emerged only during the early 1920s. (Foshay, 1969, p.275) It was not until 1918 that the first book devoted to the curriculum was published. Written by Franklin Bobbitt and titled simply 'The curriculum' this volume is recognized as the milestone that marks the emergence of curriculum as a field of study (Zais, 1976, p.5).

Schubert (1982, p.420) has stated that curriculum research goes hand in hand with curriculum scholarship. He has further observed that curriculum scholarship is more properly denoted by the terms "inquiry", "studies", "theory", and "perspectives" rather than research. The broad conception of curriculum research is sometimes empirical, but more often analytic, conceptual, critical, and/or normative.

Thus, curriculum research concerns inquiry about the course of educational experiences. It usually focuses on experiences of learners in educational institutions, and embraces the formulation, implementation, and outcomes of curricular policy.

As curriculum research contributes to a field of study and not to a discipline, there is no research methodology that it can call its own (Taylor, Reid & Holley, 1974a, p.3). It must draw on methodologies as its problems dictate and as the focus of its attention demands. These will be the methods of the philosopher, historian, literary critic or social scientist.
2.3.1 A Systems Model of Scope and Concerns of Curriculum Studies

Taylor and Colin (1979, p.20) have proposed a systems model depicting the scope and concerns of curriculum studies. This model is depicted in Figure 2.2. This systems model clearly depicts the relationship between the Process of curriculum design and development, Intended Curriculum, Operational curriculum and Curriculum evaluation. The feedback loops in the systems model involve judgments of various kinds related to the activities of curriculum design, development and transactions.

An input to the process of curriculum development is a point of view about the purpose education is to serve. The process of curriculum development starts working from aims, and develops related objectives, subject matter content, curriculum materials and a structure or sequence for what should be taught.

The output consists of a range of intended curricula which can be assessed for their congruence with the aims of education which were an input to the curriculum development process. The process of assuring congruence is the feedback from the output to the input of the curriculum development process.

Each stage in a systems operation can be studied separately, for the kinds of decisions and actions which are involved. A beginning for the major curriculum system has been made in this model (Fig.2.2), where the decisions and actions are given in terms of some of the questions which are asked in each system. Such a systems framework is not all-embracing nor can it be taken as a definitive model for curriculum studies. It does, however,
Figure 2.2 THE SCOPE AND CONCERNS OF CURRICULUM STUDIES: A SYSTEMS MODEL

APPRECIATIVE JUDGEMENT

CONCEPTIONS AND IDEOLOGIES OF EDUCATION
Transmission of cultural heritage
Education for self-realization
Education for democracy

CURRICULUM DESIGN AND DEVELOPMENT
Educational aims
Objectives of the curriculum
Subject matter (or educational experiences) selected from disciplines, fields of study, experience of educational practices.

CONGRUENCE

OPERATIONAL CURRICULA
Teaching x Learning x Content x Method of teaching and learning
Organization of teaching-learning group x milieu in which learning takes place

EVALUATION

EDUCATED INDIVIDUALS
Individuals with developed capabilities, skills and attitudes which enhance desired qualities

What is education? What purpose is it to serve? What are valid aims for it? What are the values underlying these purposes?

What should be taught? How should curricula be designed? Who should make decisions about development? How should such decisions be made? What means should be used to indicate what should be taught?

What is to be taught? What is to be learned? What standard is to be achieved? What ground is to be covered?

How are the pupils organized? What teaching methods are employed? How is the material being presented? How is the content of the curriculum divided into teaching units? Who and what influences these decisions? What are the characteristics of the 'milieu' in which teaching takes place?

Judgements of value and assumptions
Decisions and actions based on judgements of value and experience of reality
Result of the curriculum development process
Decisions and actions based on pedagogic experience
Results of curricular experience

Source: Taylor and Colin (1979, p. 20)
provide an overall framework which helps give form and coherence to this field of educational study.

Understanding the issues involved in activities concerned with curriculum development and the factors influencing the success or failure of such activities is a crucial area of curriculum studies. Without an understanding of the issues teachers remain at the mercy of events, of unnoticed assumptions, of unrecognised influences and of the prejudices of habit and practice.

Equally important to an understanding of the process of curriculum development, is an understanding of what happens to intended curricula as they are worked upon in schools and classrooms. How intended curricula are enacted, how they become operational, the factors which may affect them and result in unintended effects are all important to study.

Curriculum evaluation studies constitute a major area of curriculum studies. Curriculum evaluation is essentially concerned with judging curricula through processes of measurement or valuing or a combination of the two. In the systems model (Fig.2.2) curriculum evaluation is shown as the feedback loop from educated individuals to intended curricula and thence to the factors governing operational curricula.

In the past, curriculum evaluation was somewhat restrictively conceived, focusing over much on either the output (product) of operational curricula or on their processes to the neglect of the evaluation of what was antecedent to both, the process of curriculum development and the input to this process.
derived from conceptions and ideologies of education. In drawing attention to another aspect of evaluation, in this case, the congruence between the qualities inhering in intended curricula and the originating conceptions of education, Stake (1967) has made a singular contribution.

The work of Vickers (1968) on how social judgements are made in practice has led to a realization, as observed by Taylor and Colin (1979, p.136) that a larger form of curriculum evaluation—'appreciation' has been neglected. The notion of judgement ties both styles of evaluation viz., judging by measuring and judging by valuing together. In the past curriculum workers have placed greater emphasis on measurement.

However, more recently, the centrality of values in evaluation has been stressed. Values enter into the determination of curricular aims and objectives, into the means proposed to achieve these and into the interpretation of any measurement process that may be employed. Values are involved in the understanding, meanings, interpretations and motivations of those concerned with curriculum development and operational curricula. A particularly important aspect of evaluation is 'Congruence' or the establishment through judgement of the extent to which values embodied in conceptions of education are incorporated in intended curricula. This is a more reflective process than a measurement process. It is usually less deliberately exercised than the process of measurement and tends to be overlooked.

'Curriculum appreciation' which is an important form of evaluation takes into account both facts and values. It aims to
make a judgement about whether what is taking place is what is wanted, to consider the worthwhileness of the educational process (including the role which the curriculum plays in it) and to decide whether or not to propose changes. The term 'Appreciative judgement' used in the systems model shown in Fig.2.2. refers to valuing the qualities of the curriculum/curricular activities in relation to all forms of available evidence. It is concerned with overall appraisal and practical recommendations. National reports such as the Plowden Report attempt an overview of the educational system or part of it by examining it and considering what ought to be going on. Such reports engage in a process of appreciation, they attempt a just valuation of the process of education at the heart of which is the curriculum.

2.3.2 Tyler's Paradigm of Curriculum Research and its Categories

Prior to the 1950s, curriculum researchers built their research around a host of topics. Tyler (1949) proposed a paradigm which provided a perspective needed to give shape and cohesion to this diversity. Tyler's paradigm identifies four major curriculum topics: "Purposes", "learning experiences," "organization", and "evaluation". The four topics of the paradigm serve as an outline for portraying major categories of curriculum research. A brief description of research on the first three aspects is presented in this section; research on the fourth aspect viz., curriculum evaluation is discussed under 2.4.0.
2.3.21 Research on Curricular "Purposes"

Curricular purposes are denoted by the terms 'aims', 'intents', 'ends', 'goals', 'objectives' and 'competencies'. Curricular purpose has been studied relative to both form and substance.

Behavioural objectives is the mostly widely used form of stating curricular purposes. Arguments in favour of behavioural objectives rest on the manageability of their tangible form as compared with global statements of purpose. Criticisms of behavioural objectives and competence approaches have been made by noted researchers like Atkin, Stake, Stenhouse, and Eisner (Hamilton et al., 1977).

Determination of the substance of curricular purposes, i.e. the qualities to be fostered in students has been investigated by many researchers. Tyler (1949) called for studies of learners, contemporary life, and scholarly disciplines. This emphasis on learner, society, and knowledge is reflected in Dewey (1902), the Eight-year study (Aikin, 1942) and Taba (1962).

Bobbitt (1924) suggested that 'activity analysis' could be used to formulate curricular purposes. By studying what people do in order to perform successfully in adult life, he translated adult activities into curricular objectives. In the fields of technical and vocational education 'Job analysis' is carried out to derive curricular objectives.

2.3.22 Research on "Content or Learning Experiences"

Curriculum content is conceived differently by different authors. One conception holds to content as subject
matter derived from scholarly disciplines (King & Brownell, 1976). Another stems from early social behaviorists and equates content with activities. Tyler (1949) is often associated with either position, but he defines "learning experiences" to indicate the interdependence of content and the learning process.

Research in this area deals with the nature of content/learning experiences and the methods for its selection. According to Tyler (1949) the planning of content should be a gathering of evidence that enables curricularists to predict and create desirable changes in the experiential repertoire of students. To this end, Parker and Rubin (1966) and Berman and Roderick (1977) provide Deweyan interpretations that conceive of process as the most defensible content. Smith, Stanley and Shores (1957) have illustrated that content is derived from the following sources: Scholarly disciplines, academic subjects, student needs, social practices, universal institutions and significant social problems.

2.3.23 Research on "Curricular Organization"

Schubert (1982, p.424) has categorized the vast amount of research on curricular organization in terms of: scope and sequence; instruction and methodology; and several environmental dimensions (human, physical, material and psychosocial).

2.3.231 Research on Scope and Sequence

The scope of curricular organization refers to the range of content or experiences to be provided. Questions about
the proper balance of curricular offerings, core versus elective subjects, and depth versus breadth deal with scope.

Sequence deals with the proper ordering of content, that is, the matter of prerequisite knowledge. Research reveals two different criteria for determining sequence viz., 'psychological' and 'logical'. (Dewey, 1916). Proponents of progressive and open education have followed the psychological approach (which takes student interests or current sources of meaning in student lives as bases for sequence), whereas those who guided curriculum reform projects in the late 1960s followed the logical approach to organization. (Logical organisation proceeds according to the structures of academic disciplines). Dewey (1916) suggested that we must begin with the psychological, and help students move toward acquisition of the logical. According to Kohlberg and Mayer (1972), this position is strongly consonant with Piaget's notion of developmentally appropriate learning. Gagne (1967), however, suggests that sequence should take care of hierarchy i.e., sequence should move consecutively from simple to complex capabilities. Schubert (1982, p.424) has stated that research on curriculum practice reveals the power of the less rational sources such as social trends, ideology, political pressure, teacher preference and the marketing strategies of publishers in influencing the sequence.

2.3.232 Research on Instruction and Methodology

Some researchers prefer to separate curriculum from instruction for analytic clarity, whereas others regard
separation as superficial since curriculum and instruction are thoroughly intertwined in practice.

Broudy and Palmer (1965) have provided historical perspective on instructional methods; Travers (1973) has provided research overviews on instructional strategies; Dunkin and Biddle (1974) have summarized empirical studies of teaching, and Joyce and Weil (1980) have explicated models. Recently much attention has been given to the topics of 'Mastery learning' and 'aptitude - treatment - interaction', in which achievement is examined as an outcome of learner predisposition and educative input. As observed by Schubert (1982, p.424) current research clearly points out the interdependence of instruction and curriculum.

2.3.233 Research on Environmental Dimensions

Environmental dimensions pertain to human, physical, material and psychosocial aspects.

The grouping of human beings (Students and Teachers) is a major field of study which has long involved the issue of homogeneous versus heterogeneous grouping and the concomitant criteria for group selection. The organisation of teachers in relation to students has also been studied.

Research on physical aspects of the environment has dealt with arrangement of the physical plant like the comparative worth of openspace, modular, and pod arrangements, as well as more venerable curriculum issues such as subject, activity and core curricula. Strengths and weaknesses of the latter are discussed by Smith, Stanley and Shores (1957). Research findings have highlighted that variations on each curriculum pattern utilize
distinctive physical arrangements of furniture and equipment that markedly influence purposes, content and evaluation.

The organization of published and locally made instructional materials has been studied in its many dimensions. The Educational Products Information Exchange (EPIE) argues for greater clarification of selection and use procedures (1979). EPIE tries to enable educators to clarify their needs and select materials that are designed with internal consistency as to intent, content, methodology, and evaluation. Goodlad, Klein and Associates (1970) have observed that implementation of innovative curricular materials rarely remains consistent with intent.

The psychosocial dimensions of curricular organization is concerned with the educational atmosphere and its conduciveness to learning. Researchers of learning environments have probed psychosocial factors quantitatively. Walberg and Moos (1980) have assessed Organizational, suprapersonal, and social as well as physical aspects. Fraser (1981) has incorporated student perceptions of learning environments in curriculum decision making and evaluation. (cf.2.7.0)

2.4.0 CURRICULUM EVALUATION

Curriculum evaluation constitutes the fourth topic in Tyler's paradigm presented in the previous section (2.3.2). As observed by Schubert (1982, p.425) early in this century evaluation was nearly synonymous with testing and measurement. Between 1930 and 1960, writers of synoptic curriculum texts expanded evaluation to include a variety of evidence on student

2.4.1 Definition of Curriculum Evaluation

The evaluation group of Harvard Project Physics (Welch and Walberg, 1968) has defined curriculum evaluation as those activities which provide information useful in course improvement and which show how effective the course is under specified school conditions.

Lewy (1985, p.198 and 1977, p.30) has proposed a mapping sentence definition of curriculum evaluation which serves as a classification scheme of curriculum evaluation studies. An adapted version of this is presented in Figure 2.3.

Mapping sentences of this type have been employed for the comparison of several curriculum evaluation approaches (Lewy and Shye, 1974 and Levy, 1977).

The mapping sentence contains three facets: the stage of the curriculum development; the component of the curriculum or the entity being evaluated; and the type of decision situation. Combining all the three aspects it is possible to describe curriculum evaluation in the form of a mapping sentence.

The overall definition presented in the mapping sentence in Figure 2.3 contains a variety of evaluation activities; their totality makes up the more general concept "curriculum
evaluation". This definition suggests that evaluation is the provision of information for the sake of facilitating decision making at various stages of curriculum development. This information may pertain to the programme as a complete entity or only to some of its components.

The mapping sentence in its totality constitutes an overall inventory of decision situations which one may encounter during the process of curriculum development. Each such decision situation may require the conducting of a short-duration focused evaluation study.

**Figure 2.3 Mapping Sentence Definition of Curriculum Evaluation**

A: Stages

| Evaluation is the provision of information at the | determination of aims | planning |
| | tryout | field trial |
| | implementation | quality control |

Stage of programme development

B: Entity

| concerning | course content | instructional resources | study material | teacher's guide | methodological approaches | whole curriculum |
| | for the programme |

C: Decision Situations

| sake of making decisions about | selecting elements of | modifying | qualifying the use of |
| | the programme |

Adapted from Lewy (1985 & 1977)
One may define a particular substudy in the process of curriculum evaluation by selecting a single line from the three facets appearing in the mapping sentence. Thus, for example, one may conduct a small focused evaluation study at the field trial stage of a programme concerned with the quality of the textbook. Such a study will provide information needed for making decisions about modifying the original version of the textbook.

The definition of the present study formulated by selecting a single line from the three facets appearing in the mapping sentence definition of curriculum evaluation is given below.

The present study aims to provide information at the quality control STAGE of programme development concerning the whole curriculum in chemistry (ENTITY) for the sake of making DECISIONS about modifying the programme.

In a similar fashion one may define other substudies. As indicated above, the totality of all substudies concerned with a particular programme constitutes its evaluation. The validity of this definition is dependent upon the fact that all empirical activities can be described by some combination of the key words; however, there is no requirement that all possible combinations should correspond to some empirical phenomenon.

The mapping sentence summarizes the variety of evaluation studies that may be performed during the life cycle of any new programme. Nevertheless, one should be cautious not to over evaluate a programme. A great variety of evaluation foci have been mentioned here, not for the sake of encouraging the evaluator to utilize all of them in the context of dealing with a
single programme, but to provide a broad inventory, from which activities most relevant to answer crucial questions may be selected.

2.4.2 Models of Curriculum Evaluation

Formalizing a complex process such as curriculum evaluation into a model is very helpful. The function of a model in evaluation is to provide a conceptual framework or a rationale for designing evaluation studies.

Given the range of positions evaluator can formulate from underlying philosophical assumptions, the proliferation of models for conceptualizing curriculum evaluation can be easily understood. According to Boruch and Wortman (1979), there is no one generalized model for conducting evaluations. Variations in curriculum evaluation models result from differences in the purpose of evaluations, the types of evaluations, the methodology used in the conduct of evaluation and the questions asked.

As evaluation efforts sought curriculum improvement, researchers acknowledged the importance of process variables. Scriven (1967) distinguished between "formative" evaluation, focusing upon implementation processes, and "summative" evaluation (focusing upon outcomes). Stufflebeam (1969) described formative elements in terms of "Context", "input", and "process", but his interpretation of "product" can be associated with the summative approach. Stake (1967) added complexity to the conception of evaluation by highlighting three major variables subject to both descriptive and judgemental portrayal - "antecedents", "transations" and "outcomes". Further, he (Stake,
1975) has advocated a form of evaluation in which evaluators strive to be responsive to evolving client needs rather than to their preordained goals. Scriven (1972) has proposed a goal-free model in which evaluators portray curricular conditions without being biased by the goals that clients say they seek.

2.4.21. Classification of Models

Popham (1975) has classified evaluation models into the following three categories:

i) **Goal Attainment models:** Examples of goal attainment models are the objective-based models exemplified by the work of Tyler (1966) and Metfessel and Michael (1967).

ii) **Judgemental Models:** These models can be further subdivided into two classes in terms of intrinsic and extrinsic judgements. Accreditation, such as that established by the National Study of School Evaluation (1978) in the U.S.A. would fall into the intrinsic category of judgemental models. Scriven's (1972) "Goal-free evaluation" model and Stake's (1967) "Countenance model" can be placed in the extrinsic category of judgemental models.

iii) **Decision-facilitation Models:** The CIPP model of Stufflebeam and Guba (1970) and the discrepancy model of Provus (1971) are examples of decision-facilitation models.
2.4.3 Methodological Approaches in Curriculum Evaluation

Methodologies used in curriculum evaluation are classified along a continuum. As observed by Talmage (1982, p.603) the bipolar labels often used include soft and hard research (Cronbach and Shapiro, 1978), quantitative and qualitative research (Patton, 1980), objective and subjective epistemologies (House, 1978), conventional and naturalistic research (Guba, 1978), and behaviouristic and humanistic research. These polarities indicate proximity to or distance from the methodology of experimental research.

The methodological approaches used in curriculum evaluation can be classified into three broad categories as shown below:

i) Experimentalist approach
ii) Qualitative approach
iii) Eclectic approach.

A brief description of each approach is presented below

2.4.31 Experimentalist Approach to Curriculum Evaluation

Experimentalist evaluators focus upon establishing causal links between the programme and the outcomes. Cook and Campbell (1979), the RMC group (Horst, Tallmadge and Wood, 1975) and Rivlin and Timpane (1975) represent this position.

Much has been written about the difficulty of making meaningful comparisons between achievement levels of students in the experimental or innovative programme and those of students in the old or traditional programmes. Forehand (1966) lists a number of difficulties. The first one is the impossibility of
controlling or isolating the effective features of an experimental curriculum. Another lies in the attempt to define a control group that differs with respect – and only with respect – to the central innovative idea of the experimental programme.

Though often treated as if it were unitary, a curriculum actually involves many elements, not all of which covary systematically; reading materials, sequences of instruction, training of teachers, teaching methods, teaching manuals, and teaching aids to cite a few. Each of these elements is in fact a variable, and does in fact take on different values in curriculum variation.

2.4.3.1 Alternatives to the "Experimental Design" Approach

Where true experiments are not possible, then Cook and Campbell (1979), Forehand (1966) and others propose quasi-experimental designs such as interrupted time-series, control-series design, regression discontinuity, and multiple group pre-post comparisons.

An alternative proposed by some experimentalists is termed "natural experiments" which are also called as "Planned Variations" (Talmage, 1982, p.599). Outcomes related to variations in curricula with common goals are compared. The Follow Through programmes represent planned variations (Rivlin & Timpane, 1975).

2.4.3.2 Qualitative Approaches to Curriculum Evaluation

Qualitative approaches to curriculum evaluation, have been developed primarily in response to several critics' observations that evaluators have been preoccupied with whether student
outcomes match prespecified objectives. These critics have stated that students acquire other than solely intended learnings that must also be evaluated, and that an examination of the entire process of education, not only its outcomes, is necessary inorder to improve curricula. Both the overt and hidden curriculum can be evaluated by means of qualitative curriculum evaluation. The qualitative approach provides a deep, complex understanding of the curriculum in use. It presents a holistic portrayal of the curriculum in action, and projects the perspective of the participants.

McCutcheon (1982, p.1503-1506) has presented an analytic description of eight of the most commonly used approaches of qualitative curriculum evaluation. The biographical approach (Berk, 1980 and Grumet, 1980) aims to convey the impact of curricular activities upon an individual. Case studies (MacDonald & Walker, 1977) document and reveal how a curriculum is interpreted in its use in varied settings. Educational critics (Eisner 1979 and McCutcheon, 1979) who are connoisseurs present the nature of a curriculum in use through description, interpretation and appraisal. Ethnographic approach (Smith and Keith, 1971) reveals the extent to which a particular curriculum fits with the wider culture. Illuminative evaluation (Parlett & Hamilton, 1977) reveals the process of curriculum implementation. Portrayal evaluation (Fraser, 1980 and Kemmis, 1977) creates a vicarious experience so that the audience can be informed and make judgements. Responsive evaluation (Stake, 1975) characterises programme activities with respect to the audience's
requests for information and varied value perspectives. Portrayal, responsive and illuminative evaluations are generally considered to be phenomenological, as they reveal what it is like to be a teacher or student using the curriculum; the felt meaning of the curriculum, along with the values and feelings of participants. Critical science evaluation (Apple, 1979) attempts to interpret curricular phenomena through a marxist perspective.

2.4.33 Eclectic Approach to Curriculum Evaluation

As described by Bryk (1978) eclectics attempt to draw on the strengths of each approach - the objectivity of quantitative data and the richness of qualitative information to create an integrated view of programme impact. Eclectic approach is based on a pragmatic philosophy. This position is most prominently represented by Bryk (1978), Cronbach and Associates (1980) and Weiss and Rein (1972). These evaluators take an eclectic position by drawing upon experimental designs (mostly they use quasi experimental designs) that make possible causal statements, along with Qualitative methods that describe the process of curriculum implementation as well as contextual variables effecting the curriculum. In this way an evaluation design can search for multiple causality or generate plausible explanations that approximate reality.

Cronbach et al. (1980, p.233) sum up the eclectic position: "The evaluator will be wise not to declare allegiance to either quantitative - manipulative - summative methodology or a qualitative - naturalistic - descriptive methodology. He can
2.4.4 REVIEW OF CURRICULUM EVALUATION STUDIES

In the first part of this section some curriculum evaluation studies carried out in other countries have been reviewed and the second part presents a review of the studies carried out in India.

2.4.41 Curriculum Evaluation Studies Carried out in other Countries

In the early 1960s national curriculum study projects such as the Physical Science Study Committee (PSSC) and Chemistry Educational Materials Study (CHEM Study) and Biological Science Curriculum Study (BSCS) at the senior high school level employed with some variations the evaluation model and procedures advocated by Tyler (1949).

While reviewing the studies on curriculum evaluation one frequently encounters comparison with control groups either in a randomly selected experimental setting or in a nonrandomly assigned quasi-experimental setting. For example the Harvard Project Physics (Welch, 1973) employed a randomly assigned control group design in their evaluation. Multiple measures were employed at all levels of analysis — learning outcomes (both cognitive and affective), student entering characteristics, teacher attributes, analysis of the learning environment and both student and teacher reactions to the course. In this study teacher variables as teacher's knowledge of Physics, teacher
personality and teacher attitudes towards Science and Teaching, generally bore no relation to student achievement. (cf.2.7.711 and 2.7.72)

The studies conducted by Weber and Renner (1972) showed that the Science Curriculum Improvement Study Group had significantly greater gains in utilizing science processes than a matched group using a textbook.

Furst (1950) used the technique of factor analysis for studying the effect of the organisation of learning experiences upon the organisation of learning outcomes. Reid (1978) has also employed factor analysis to identify the constraints on teaching school curricula.

Intensive classroom observation and the development of ways to classify teacher student interaction (Medley and Mitzel, 1963) result in useful information for curriculum evaluation. Such classroom visitations help to determine if teachers are pursuing the program, as outlined by course developers. Shulman and Tamir, (1973) have stated that measures of student perception of a learning environment can serve as useful adjuncts to, if not substitutes for, direct observation.

Raths (1978, p.245) pleads for identification of variables, that can be manipulated by teachers and administrators. He is of the opinion that mere focussing on variables of sex and race and Socio Economic Status (SES) levels, though may be important in the inquiry into a curriculum's effectiveness, will not serve the purpose. This is because nothing can change a student's SES level, if that variable is found to be correlated with
achievement within a programme. He argues that curriculum evaluators must examine instructional variables such as the extent to which teachers give emphasis to one area of content over another; the degree to which students are given the opportunity to practice what they are learning; and the quality of the feedback that students receive about their work. What makes these variables so important is that as they are identified as keys to the teaching-learning situation, they can be manipulated by teachers to increase the effectiveness of the curriculum.

Based on eleven case studies of District Science instruction in Kindergarten through Grade 12, Stake (1979) found that a lack of communication existed between the District Curriculum Coordinators and testing personnel, that curriculum coordinators paid little attention to test results, and that testing people were not concerned about curriculum review. In another case study they found no instances of curriculum change or recommendations for change based directly on student testing. Stake is of the opinion that using group test means as a basis for curriculum diagnosis or programme evaluations is a procedure which is not grounded in validity studies, but is based on experience and intuition.

Prindiville (1975) has reported the findings of a Civil Highway curriculum evaluation carried by Northeast Wisconsin Vocational, Technical and Adult Education District, Green Bay. The purpose of the evaluation study was to provide data for course review and revision. As civil engineering is a rapidly changing occupational field, the study aimed to determine the
skills that are currently essential for civil highway engineering. The curriculum of the associate degree programme in Civil Highway engineering was summarized and presented to both programme graduates (1970-1973) and to potential and actual employers for their assessment of the importance of the abilities required on the job. The principal conclusion from the data was that more emphasis is needed on communication skills, particularly written reports. Conflicting responses in the areas of mathematics and sciences indicate that further research is needed in these areas.

Gerlovich (1980) has reported the development of a tool for assessing and revising science curriculum. The tool was jointly developed by Iowa Department of Public Instruction and a committee of the Iowa Council of Science Supervisors to encourage and aid local schools in assessing their science curriculum on a continuous basis.

The Division of Science Education of the State Department of Public Instruction, (1975) of North Carolina State in the U.S.A. has developed a "Science Education Assessment Instrument" for individual or group use for assessing science curricula. Specific items within seven broad topics (Foundations for local planning, curriculum, teaching-learning, staff, facilities, equipment and materials, evaluation, and finances) are scaled from zero (totally lacking) to four (excellent). The scale for each topic is averaged and plotted on a profile chart. It also includes a plan of action form for planning science curriculum improvement based upon the areas of strength and need identified on the profile chart.
The Nuffield Advanced Chemistry Research Project has been evaluated by Leece and Mathews (1976). Besides concentrating on the uses of examination outcomes, they have looked into other sources of information for a wider perspective: these include questionnaires to teachers and students about the curriculum and examination and the attainment of subgroups of students such as those studying mathematics compared with those who do not. In this study teachers and students perception of the facility of each curriculum topic (related to the average level of difficulty of the whole course) was collected and compared. The teacher and student estimates agreed closely (rank correlation 0.83). However, neither teacher nor student estimate agrees closely with the estimate of examination facilities of curriculum topics. The students relative interest in each curriculum topic was also estimated using a five point scale and analysed. Two interesting findings of the study are (1) Mathematics students find in a broad sense, to be better chemists (as judged by the Nuffield examination) than non-mathematicians, and (ii) the overall performance of girls is slightly superior than that of boys.

Tamir (1985) in his recent book which presents an excellent overview of the evolution of curriculum evaluation since the reform of the 1960s has edited and presented eight case studies. Seven out of these eight case studies, deal with the evaluation of science curricula. These represent diverse educational systems, in a range of countries including Australia, Israel, England and the U.S.A. In this book (Tamir, 1985) Hulda Grobman has reflected her experiences as the first in-house evaluator of
the BSCS. Fraser has described the formative evaluation of the Australian Science Education Project. Bond, Dynan, Pasker and Ryan have described how they utilized the illuminative model of evaluation for evaluating an innovative physical science course in Western Australia. Two different accounts of evaluating the same curricula viz., Scottish Integrated Science course has also been presented. The first, by Kellington and Mitchell, represents a rather traditional approach which seeks to find out to what extent the course objectives have been attained. The second, by Brown presents a relatively novel approach to evaluation in which the evaluator attempts to uncover the hidden agenda of curriculum development and implementation. The importance of conceptualizing evaluation findings under an explanatory system is illustrated by Elliot. The last two case studies deal with two projects in Israel. Tamir has described the evaluation of Israeli High School Biology Project as an example of ongoing evaluation of a curriculum development project affiliated with a university. Lewy has described the mode of operation and the roles played by an evaluation unit which is responsible for evaluation of many projects in the Israel curriculum centre. Thus a variety of innovative curricula are focussed by these eight case studies and the models which emerge are empirically based. Their diversity provides evidence for the need to accommodate and adjust theoretical and methodological principles to real situations.
2.4.41 Syllabus Analysis

Feisel and Schmitz (1979) have prepared a curriculum diagram in electrical engineering by following the technique of curriculum analysis. Curriculum analysis which is also known as syllabus analysis is the process of breaking a course or curriculum into its component parts and identifying prerequisite relationship. Wyant (1973) has employed the technique of Network analysis to analyse the syllabus of the ordinary National Diploma Course in Technology conducted at Coventry Technical College. The drawing of the network closely followed the discipline imposed by the technique which asks "What is the earliest time I could possibly teach this subject or topic?" and also "What is the latest possible time that I must have taught it by?"

Zais (1976, p.383-388) has highlighted the need for evaluating the coherence of curriculum elements. He has stated that the effectiveness of a curriculum depends to a large degree on the coherence of its inter related components.

2.4.42 Curriculum Evaluation Studies Carried out in India

In India, research on curriculum evaluation is still in its infancy. Even in the Second, Survey of Research in Education (Butch, 1979) which presents a review of research at Ph.D and project levels carried out in India during the period from 1972 to 1978, there is no separate section on curriculum evaluation. The chapter on "Curriculum, Methods and Textbooks" (Roy, 1979) also does not contain a separate section on curriculum evaluation studies. Under various subjects, a few curriculum evaluation studies have been mentioned; most of these studies deal with the
curricular content and methods of teaching of English and other Indian languages. As observed by Roy, (1979, p.283) "In respect of components of the curriculum, there was a definite lack of attention on knowledge subjects such as Mathematics, Science, History, Geography, Social studies etc".

Gothiverekar (1947) first made a comprehensive study of the secondary school curriculum in the province of Bombay. He recommended a new curriculum catering to all aspects of the development of the learner. Chanana (1967) made a historical survey of the high school curriculum in the Punjab during the twentieth century and advocated a new and effective secondary curriculum. Pillai (1968) investigated into the changes in the content and scope of the primary as well as secondary school curriculum in Kerala during the thirty years since 1934. After examining the syllabi, question papers, textbooks, administration reports and reports of the expert committees he came to a conclusion that although in the light of aims and outcomes the curriculum fulfilled the basic requirements as far as its contents were concerned, much more remained to be done for raising the standard of education.

Srivastava (1968) conducted a study to identify the important Characteristics of achievement of student's different areas of curricular learning and to study the effect of intelligence and sex on the achievement of students in different areas. He has employed factor analysis to identify the functional relationship between the areas. In "Science" the following three factors have been identified. viz., 'Numerical
Factor', 'Factor much needed for grasping principles' and 'Mastery of Facts'.

Wanchoo and Sharma (1974) who surveyed the researches conducted in science and mathematics education in India found that research work done in the area of evaluation was mostly confined to test construction.

Shah (1975) made a critical inquiry into the progress of home science education in the secondary schools of India, while Deulkar (1967) evaluated home science curriculum with special reference to its functional implementation and the personal and professional satisfaction of the students.

Gupta (1977) has conducted an experimental study for upgrading the Science syllabus of Grade 7.

In the field of evaluating instructional materials, National Council of Educational Research and Training (1972) has developed evaluative criteria for assessing textbooks.

Technical Teachers' Training Institute (Western Region (1977)) has carried out an evaluation of the curricula of Diploma courses in Engineering offered by the polytechnics in Madhya Pradesh State. The evaluation covered the course content, teaching and learning strategies, adequacy of instructional time, teacher and student support materials, resources of the polytechnics, the communication system in the polytechnics and the examination system. A similar study has been conducted by the same institute (Technical Teachers' Training Institute, (Western Region) 1982) for evaluating the curriculum of the polytechnics in Gujarat State. In both the studies, information has been collected from
the present and past students, teachers and employers through questionnaires and interviews. On site visits to the polytechnics for observation has also been made.

Brahadeeswaran, Natarajan and Kanakaraj (1983) have conducted a study to evaluate the curricula of Diploma courses in Engineering offered by the Polytechnics in Andhra Pradesh State. Evaluative information collected from students and teachers (through questionnaires) was analysed and reported to facilitate curriculum improvement.

To sum up the review of curriculum evaluation studies in India the following inferences can be made.

* Among the limited number of curriculum evaluation studies reported, most studies pertain to curricula of English and other Indian languages.

* Very few studies deal with the curricula of Science subjects.

* During the past ten years Technical Teachers' Training Institutes have conducted a few studies to evaluate the curricula of polytechnic courses.

* The studies under review show that experimental designs with the application of rigorous statistical analysis, are rare. Quite a large number of studies are survey type. Achievement tests, questionnaires and interview schedules are the most frequently used tools for data collection.
2.5.0 ASSESSING THE EFFECTIVENESS OF CURRICULA

Curriculum's effectiveness is not easy to determine because the curriculum is complex, evaluators have different viewpoints, and the curriculum is intertwined with other elements of the educational institution. Questions about curriculum's effectiveness are further complicated by time. A student who expresses satisfaction with the curriculum during the period of his studentship may find upon retrospect as an adult, that the curriculum of the course he underwent was deficient. The reverse may also happen. The student who is discontent in school may discover that his adult achievements were enhanced considerably by the schooling he received.

The best way to make sense out of answers about curriculum effectiveness is to identify the kind of evidence required to decide its effectiveness. When this evidence is specified, effectiveness can be judged according to whether the curriculum meets the established criteria. Thus effectiveness is always referred to some criterion measure specified.

According to Herald (1982, p.10) educational effectiveness is a measure of how well we are able to do that which we set out to do. Efficiency on the other hand is related to the various kinds of costs (money/time/space) associated with our efforts to be effective. In other words efficiency refers to the capacity to produce effective results relative to the efforts and resources expended.

Orlosky and Smith (1978, p.419-420) have emphasised that students' achievement is a primary factor in determining curriculum effectiveness.
Davis, Allexander and Yelon (1974, p.11), point out that the best measure of instructional effectiveness is the number (percentage) of students achieving the course objectives.

Walberg's multi-factor psychological theory of educational productivity holds (1980) that learning is a multiplicative, diminishing-returns function of student age, ability and motivation; of quality and quantity of instruction; and of the psycho-social environments of the home and the classroom.

Gormly (1981, p.70) advocates comprehensive measurement for measuring effectiveness. He uses the term comprehensive assessment to refer to a strategy of assessment design in which we use multiple perspectives rather than a single index of performance such as academic attainment.

Hartnett and Centra (1977) attempted to relate student performance in different academic fields to various characteristics of the academic faculties in which the students studied. Criterion measures of achievement tests showed large differences in the apparent "effectiveness" of the different departments, which were selected from a number of different universities. Students' pre-entry levels of achievement were controlled. Analysis of faculty - student ratios, faculty interest in teaching, faculty salaries and department size failed to demonstrate features consistently associated with effectiveness. The investigators suggest that students' perceptions may be more important in explaining effectiveness; indeed, Centra (1976), among others, has shown that student ratings of teaching effectiveness are positively related to mean student achievement.
Stork (1981) conducted a study to design and field-test a model for determining the effectiveness of mathematics instructional programmes within the California community college system. Multiple criteria for the evaluation model were developed by means of an examination of related literature and input gathered from administrators, teachers and students. Criteria related to the cost effectiveness of the mathematics programme was considered by those surveyed as inappropriate for determining instructional effectiveness.

Marco, Murphy and Quirk (1976) have presented a classification of methods of using student data to assess the effectiveness of schools. Their classification is based on two major dimensions: (a) the type of data used for assessing performance (e.g. one shot data with or without input data, cross sectional data and longitudinal data.) and (b) the type or referencing used for setting standards viz., Norm referencing/ Criterion referencing.

McGuigan and Peters (1965) have reported an investigation of the suitability of different measures of pupil achievement for evaluating programmed materials. They suggested 'gain as a proportion of possible gain'. This was termed 'gain ratio' and an arbitrary value of 0.50 was set as a criterion by which good programmes were distinguished from those 'unsuitable for publication'.

\[
\text{McGuigan Gain Ratio} = \frac{\bar{y} - \bar{x}}{T - \bar{x}}
\]
Where \( \bar{x} \) = mean group pretest score
\( \bar{y} \) = mean group post-test score
\( T \) = maximum possible test score

Bolton (1972) has investigated the effectiveness of part of an undergraduate curriculum in Electrical Engineering. The study shows that tests of cognitive abilities can be used as instruments of curriculum evaluation, particularly for diagnostic purposes. He has employed Factor analysis (a principal component solution followed by varimax rotation) to identify the factors common to the cognitive abilities test and the course examination.

2.6.0 USE OF CRITERION REFERENCED TESTS IN ASSESSING THE EFFECTIVENESS OF CURRICULA

One of the most commonly used measures of curriculum effectiveness is the student test score.

Stake (1979) argues that achievement test means, based on group scores or other district wide summary scores do not provide valid measures for diagnosing curriculum weaknesses nor for initiating curriculum changes.

According to Cronbach (1963) "to agglomerate many types of post-course performance into a single score is a mistake, because failure to achieve one objective is masked by success in another direction". Moreover, since a composite score embodies (and usually conceals) judgments about the importance of the various outcomes, only a report that treats the outcomes separately will be useful to curriculum evaluators.

Glaser (1963) and Popham and Husek (1969) were the first to introduce the concept of criterion referenced testing. Their
aim was to provide the kind of test score information needed to make a variety of individual and programmatic decisions arising in objectives-based instructional programmes. The more traditional norm-referenced tests were considered less than ideal for providing the desired kind of test score information.

According to Popham (1981, p.26) "a norm-referenced test is used to ascertain an individual’s status with respect to the performance of other individuals on that test." The emphasis of norm referenced tests is on the relative interpretation, that is, the interpretation of an examinee’s performance in relation to the performance of the examinees in the normative sample. The interpretations are made absolutely for criterion referenced tests.

2.6.1 SHORTCOMINGS OF NORM REFERENCED TESTS FOR CURRICULUM EVALUATION

For several obvious reasons researchers like Hambleton and Eignor (1978) consider norm referenced tests unsuitable for the measurement of curriculum effectiveness. Because what is at issue is not whether one student’s test performance is better than another’s, but how well the student has mastered each of the objectives of the curriculum.

Popham (1978) has identified three weaknesses of norm referenced tests for programme evaluation purposes. They are listed below:

(i) Since norm referenced tests are often so general, they frequently fail to mesh satisfactorily with the curricular emphases of the programme being evaluated.
When the match between test content and programme content is low, we have nothing of value.

(ii) As norm referenced tests are very general they do not provide specific cues for identifying the weak areas of the curriculum or for instructional amelioration. The typical diffuseness of norm referenced tests renders them largely useless for such improvement guidance.

(iii) The technical item-production and item-refinement procedures employed in the development of norm referenced tests tend to make such tests less sensitive to detecting instructional effects than their criterion referenced counterparts.

The purpose of norm referenced tests is to compare an individual's performance to that of some reference group. Consequently, norm referenced tests consist of test items that contribute most to maximizing test score variability. In the process of choosing items that contribute variability, those contributing low variability are eliminated. It is clear that items tapping concepts taught successfully by a great number of teachers will contribute little to test score variability (most students will answer the items correctly) and will be eliminated, while the items measuring pure reasoning ability will have greater variability and will be retained. As a result of the process, the test begins to look less like an achievement test and more like an aptitude test. The process of item selection puts a distance between the curriculum of the educational programme and the tool used to evaluate it. The test would be
sensitive to the aptitude of the individuals rather than the effectiveness of the instruction. If an instrument is to be sensitive to the learning process, its content must be vary carefully matched to that of the programme. It is being said more and more (Hambleton and Eignor, 1978, p.15) that norm-referenced tests function like IQ tests.

2.6.2 Definition of Criterion Referenced Tests

A recent content analysis (Gray, 1978) of 57 descriptions of criterion referencing revealed that it was not unusual for different authors to use the term differently. Nitko, (1983, p.446) has observed that criterion referencing is in a state of development and the more or less standard ways of criterion referencing have not become firmly established.

Popham's (1975, p.130) definition of criterion referenced test, given below has been preferred by many researchers (e.g. Hambleton, Swaminathan, Algina and Coulson 1978, p.2).

"A criterion-referenced test is used to ascertain an individual's status with respect to a well defined behaviour domain".

The term "Criterion" in the phrase "Criterion referenced test" refers to a behaviour domain.

One of the major sources of confusion that prevailed during the last 20 years in the definition of criterion referenced test is over the word 'criterion'.

Popham (1981, p.27) has clarified this confusion by distinguishing between two conceptions of criterion: viz., criterion-as-a-level conception and criterion-as-a-desired behaviour concep-
tion. When the term 'criterion' is used to signify a desired level of proficiency, it reflects criterion-as-a-level. When the term 'Criterion' is used to signify the target behaviours themselves it reflects criterion-as-a-desired-behaviour. As observed by Popham (1981, p.28) interpreting criterion as a level of examinee proficiency yields almost no dividends over traditional testing practices. In fact, by using that conception of criterion, one could magically transform any norm referenced test into a criterion referenced test merely by setting a specific proficiency level for the test. Criterion referenced tests will provide substantial educational payoff only if they provide a precise description of an examinee's status with respect to a clearly delimited domain of behaviours.

2.6.3 CONCEPTS CLOSELY RELATED TO CRITERION REFERENCED TESTING

2.6.31 Domain-referenced Tests

Popham's (1975) definition of criterion referenced test presented in this section is equivalent to the definition of Domain-referenced test proposed by Hively, Maxwell, Rabehl, Sension and Lundin (1973). In their approach also, an examinee's performance is referenced to a well defined domain of learner behaviours.

For a variety of reasons, however, testing specialists (e.g. Popham, 1978.; Hambleton et al. 1978) have opted for the expression criterion-referenced measurement over domain-referenced measurement.
2.6.32 Objectives-referenced Tests

Objectives-referenced tests are those whose items have been constructed to measure an instructional objective. Usually such objectives are formulated behaviouraly.

As observed by Nitko (1983, p.456) objectives-referenced tests may or may not be criterion referenced. If objectives are written to describe a domain, and if items are then written to sample the behaviours in this domain, then this would fit the description of criterion referenced tests. If objectives are stated very briefly they may not provide a clear description of the behavioural domain. Poorly articulated behavioural objectives characterize what are called "ill-defined domain" (Nitko, 1980, p.466) and are not considered to lead to appropriate criterion referenced interpretation.

2.6.33 Mastery Testing

According to Nitko (1983, p.457) a test used to provide information to make a decision about whether a particular student has "mastered" a given instructional goal is called a mastery test. The term 'mastery' is likely to mean different things in different contexts. Glaser and Nitko (1971, p.641) have stated that 'mastery' in the instructional context implies that "an examinee makes a sufficient number of correct responses on the sample of test items presented to him in order to support the generalization (from this sample of items to the domain or universe of items implied by an instructional objective) that he has attained the desired pre-specified degree of proficiency with respect to the domain".
This definition of mastery is closely associated with the idea of criterion referencing. However, a mastery test need not be a criterion referenced test, although answering the question "mastery of what?" would be difficult without linking the answer to a well-defined performance domain. Given the present state of criterion-referencing, the "mastery of what" question seems to be best answered by clearly specifying the domain of instructionally relevant behaviours which a learner commands.

2.6.4 TWO BROAD CATEGORIES OF CRITERION REFERENCED TESTS

Nitko (1983, p.446-447) has distinguished between two broad categories of criterion referenced tests: those based on well defined but ordered behaviour domains and those based on well defined but unordered domains. This distinction stems from the notion that in some cases behaviors in the domain can be ordered along an achievement continuum. Not all behaviour domains can be ordered, however. Narrow domains and domains that are very homogeneous contain items that are essentially interchangeable, so that ordering is not possible. Most of the literature on criterion referenced testing have dealt with unordered domains and most of the available tests also belong to this category of 'well defined but unordered domains'.

As the Curriculum in chemistry to be evaluated in the present study consists of unordered domains, the criterion referenced tests developed for use in the present study will be based on well-defined but unordered behaviour domains.
2.6.5 THE CONCEPT OF BEHAVIOUR DOMAIN

The dictionary meaning of the term 'domain' is scope, field or province of thought or action.

When referring to "domains" in connection with criterion referenced measurement, some educators may confuse this more recent application with the former taxonomy of educational objectives context. In criterion referenced testing the term 'domain' refers to a much smaller class of behaviours (Popham, 1975, p.131).

Hively et al. (1973) have stated that the concept of domain includes both (a) specific content area as well as (b) behaviours associated with this content.

A test developer defines the behavioural domain in one of three ways: well, poorly or not at all. Tests with poorly defined or undefined performance domains do not qualify as criterion referenced tests. A well defined domain of performance is a requirement for criterion-referencing.

A domain is well-defined when both the person(s) developing the test and the person(s) using the test are clear about which categories of performance (or which kinds of tasks) are and which are not potential test items. Since the basic idea of criterion-referencing is to generalize from the few items that happen to be on the test to the broader domain of performance from which the test items were sampled, a well defined domain is a necessary condition for criterion-referencing.
2.6.6 DELINEATING THE BEHAVIOUR DOMAIN

Nitko (1983, p.451-454) has suggested the following four bases for delineating the behaviour domain of criterion-referenced tests based on well-defined but unordered domains.

(i) Stimulus properties of the domain and the sampling plan of the test
(ii) Verbal statements of stimuli and responses in domain
(iii) Diagnostic categories of performance, and
(iv) Abstractions, traits or constructs.

A brief description of each of these four bases is presented in this section.

2.6.6.1 Domain Delineation Focussed on Stimulus Properties

The stimulus properties of the items in the domain are defined in great detail. The term 'Stimulus properties' in this context refers to the surface features of the items in the domain which the test developer believes will alter the probability that a given examinee will respond correctly to a given item. After identifying the stimulus properties, attempt is made to identify those properties which may affect performance. This information is then used to form a plan for sampling items from the domain.

An early approach in this area was to define the content and content strata very precisely. (Example: Ebel's Content - Standard English Vocabulary Test (1962).)

Another more precise method was proposed by Hively et al. (1973). The stimulus dimensions of the items are used to develop item forms, which are sort of "generic" versions of items. An item form specifies an unchanging part of an item along with (a)
variable parts, (b) elements that are used in these variable parts; and (c) rules for selecting the elements.

2.6.62 Domain Delineation using verbal statements of stimuli and responses

As observed by Nitko (1983, p.453) most people associate criterion referenced tests with this category. Most people believe that behavioural objectives or some variant of them are necessary for criterion referenced testing.

Behavioural objectives with or without the cut-off score ("criterion") specified have been used to delineate the domain in the case of many criterion referenced tests. Examples: Tests based on Mager's type of objectives (1962), Curriculum Embedded Tests of IPI Mathematics (Cox and Boston, 1967), and Popham and Husek's criterion Referenced tests (1969).

"Amplified Objectives" (Popham, 1972) are considered to be more useful for preparing criterion referenced tests than the behavioural objectives. Similarly Hively et al (1973, p.13) have suggested that "Operationally defining" the objectives is a useful technique of delineating the domain. While amplifying or operationally defining an objective one has to (i) clarify the content and ability specified by the objective, (ii) provide boundary specifications regarding testing situation, response alternatives and criterion of correctness. IOX objectives-Based Tests (Popham, 1972) have been constructed on the basis of 'Amplified objectives'.

In the context of the present study "Operationally defined" objectives will be very useful as the basis for delineating the behaviour domain.
2.6.63 Domain Delineation using "Diagnostic" categories of Performance

For criterion referenced tests which are designed to make "diagnostic interpretation" especially for decisions about remedial instruction "diagnostic" categories of performance are used to delineate the domain. For example Hunt and Kirk's Tests of School Readiness (1974) are based on identifying entry level behaviours; Tests built on Resnick's component analysis (Resnick, Wang & Kaplan, 1973) are based on identifying behaviour components missing from a complex performance.

2.6.64 Defining the Domain by Abstractions, Traits or Constructs

Certain tests define the domain in terms of abstractions, traits or constructs as well as by more fine-grain behavioral objectives. As pointed out by Nitko (1983, p.454) The categories in the Taxonomy of Educational objectives (Bloom et al. 1956) refer mainly to internal processes or psychological constructs. Certain basic skills like reading comprehension and spelling ability are other examples of constructs or traits.

If the domain definition is vague (ill-defined) then they would not be called criterion referenced tests.

2.6.7 Establishing Cut-off Score for Mastery Decisions

Even though there is nothing inherent in a criterion-referenced test which requires a cut-off score, if a test is to be used to decide mastery, it is necessary to establish a cut-off score or passing standard.
The setting of a cut-off score is a validity related issue. Where the cut score is set on a test, directly affects the validity of the test because of the classification of examinees as masters or non-masters.

Berk (1985, p.1116) has stated that despite the development of more than 20 different methods, standard setting is still the stickiest technical topic. He has formulated a bilevel classification of the various methods that are used for establishing cut-off score for mastery decisions. The first level partitions the methods into two major categories based on their assumptions about the acquisition of the underlying trait or ability: State models and Continuum models. The second level classifies the methods according to whether they are based purely on judgement or incorporate both judgemental and empirical information: Judgemental methods and Judgemental - empirical methods.

State models assume that mastery is an all or nothing state; the standard is set at 100 percent.

Continuum models assume that mastery is a continuously distributed ability that can be viewed as an interval on a continuum, that is, an area at the upper end of the continuum circumscribes the boundaries for mastery. According to Berk (1985, p.1116) this conceptualization appears to fit the design and intent of most criterion referenced tests. In the present study also 'mastery' will be treated as defined by the continuum models.

As observed by Nitko (1983) and Berk (1985) a completely objective, scientifically precise method of establishing cut-off
score for mastery decisions does not exist. Regardless of how complex and technically sophisticated a method might be, judgement plays a role in the determination of the cut-off score and in the estimation of classification error rates.

The customary use of one "blanket mastery standard" for all objectives is inadvisable for the following reasons: (Berk, 1985, p.1117)

(i) Objectives vary in level of complexity and therefore applying the same standard to different objectives will be insensitive to that variability.

(ii) The time allotted for instruction (and the emphasis given in the curriculum) for different objectives will be different.

According to Pilliner (1979, p.45) a factor to be taken into account in fixing the cut-off score is the effect on future learning or, more generally, the educational consequences. If the level is set too low, pupils may be exposed subsequently to concepts or skills for which they are not yet ready. Setting it too high may hold back pupils from materials they are in fact equipped to master.

Where information about educational consequences is lacking, the guidelines suggested by Garvin (197) are useful. If on the basis of a logical analysis of the subject matter and the extant instructional system, the knowledge and skills are seen as fundamental or prerequisite to future learning, then a high proficiency level is required. A low passing score can be tolerated when the material is not seen as completing a necessary link in the development of some more complex concept or skill.
Also to be taken into account are the psychological effects on the pupil. (Pilliner, 1979, p.46) The consequences of fixing too high a passing score may be boredom, loss of motivation and damage to self-concept in pupils who do not reach it. Too low a passing score may produce psychological confusion in pupils who are moved too rapidly through the curriculum.

2.6.8 RELIABILITY OF CRITERION REFERENCED TESTS

Perhaps the first discussion of the reliability of criterion referenced tests was by Popham and Husek (1969). These authors opined that although internal consistency and temporal stability may be important characteristics of test scores that result from criterion referenced measurement, the coefficients prescribed by classical test theory for assessing these characteristics may be inappropriate. They noted the well-known result that test score reliability for a group of examinees is dependent on test score variability. Since it is not uncommon to observe rather homogeneous distributions of criterion referenced test scores, they opined that estimating the reliability of criterion referenced tests using the standard formulae applied for norm referenced tests will be inappropriate. A number of different definitions and indices of reliability for criterion referenced tests have been proposed by researchers in an attempt to cope with possible lack of score variability that attenuates traditional reliability coefficients.

One of the first suggestions for an approach to the reliability of criterion referenced tests came from Livingston (1972).
He began his work by assuming that the purpose of a criterion referenced test was to discriminate each examinee's estimated domain score from a cut-off score. It is then possible to redefine variations in estimated domain scores and domain scores about the cut-off score, rather than define the mean domain score, which is the procedure in classical test theory. Briefly, he advocates measuring score deviations from the pre-determined cut-off score instead of from the average of all scores. The farther group mean domain score is from the cut-off score, the more reliable the scores are said to be.

Livingston's procedure gives a determinate result even in the case where all testees have obtained the same score, so that score variance is zero; in this case the application of classical test theory in estimating the test reliability would have led to an indeterminate result.

However, Livingston's procedure is not without its critics. Harris (1972) points out that the standard error of measurement (which, when all is said and done, is the key statistic in reliability estimation) comes out the same whether derived by application of straight-forward classical theory or by Livingston's modification of it.

Hambleton et al. (1978) in their review of technical issues and developments of criterion referenced testing have presented the discussion on reliability of criterion referenced tests in two parts, based on the two purposes of criterion referenced tests viz. estimation of domain scores and allocation of
examinees to mastery states. In the present study the criterion referenced tests will be used for the latter purpose i.e. to classify the examinees into masters/non-masters. Hence this discussion is restricted to reliability of mastery classification decisions.

2.6.81 Reliability of Mastery Classification Decisions

Hambleton and Novick (1973) suggested that an index of reliability of criterion referenced tests which are used for mastery classification decisions should reflect the degree to which students are consistently assigned to the same mastery states across parallel test administrations, as measured by some coefficient of agreement across testings.

Suppose examinees are to be classified into \( m \) mastery states, the index of reliability tentatively suggested by Hambleton and Novick (1973) is

\[
P_O = \sum_{k=1}^{m} P_{kk}
\]

where \( P_{kk} \) is the proportion of examinees classified in the \( k \)th mastery state on the two administrations. The index \( P_O \) then is the observed proportion of decisions that are in agreement. The \( P_O \) statistic has considerable intuitive appeal and is certainly easy to calculate, but it suffers from at least one limitation.

Swaminathan, Hambleton, and Algina (1974) argued that \( P_O \) does not take into account the proportion of agreement that occurs by chance alone, and that therefore it could give a false impression to users of the extent of mastery classification
consistency. They suggested using Kappa coefficient, $k$ (Cohen, 1960) as an index of reliability. This coefficient is defined as

$$k = \frac{(P_o - P_c)}{(1 - P_c)},$$

where

$$P_c = \sum_{k=1}^{m} P_k \cdot P_k^*$$

The symbols $P_k$ and $P_k^*$ represent the proportions of examinees assigned to mastery state $k$ on the first and second administrations, respectively. The symbol $P_c$ represents the proportion of agreement that would occur even if the classifications based on the two administrations were statistically independent. Thus, in a sense, it can be argued that $k$ takes into account the composition of the group, and that in this sense it is more group independent than the simple proportion of agreement statistic, $P_o$.

The properties of $k$ have been discussed in detail by Cohen (1960, 1968). For present purposes it is sufficient to note that the upper limit is $+1$ and can occur only when the marginal proportions for different administrations are equal. The lower limit is close to $-1$. The precise lower limit of $k$ is unimportant in the context of criterion referenced testing, since any negative value indicates inconsistency and, therefore, unreliable decisions.

The coefficient $k$ is dependent upon all factors that affect the decision-making procedure: the cut-off score, the heterogeneity of the group of examinees, and the method of assigning examinees to mastery states.
The coefficient $k$ and $P_Q$ are defined in terms of repeated testings. Huynh (1976) developed a procedure for estimating $k$ on the basis of a single testing. His solution is workable, although the computation required to obtain $k$ can be tedious when there is a moderate number of possible test scores above the cut-off score.

Alternative procedures for estimating reliability from a single administration have been provided by Subkoviak (1976) who prefers to work with $P_Q$. In the present study, the single administration group coefficient of agreement proposed by Subkoviak, will be used to estimate the reliability of the criterion referenced tests. A brief description of Subkoviak's Group coefficient of Agreement is given below.

### 2.6.82 Subkoviak's Group Coefficient of Agreement

Subkoviak (1976, p.267) defined the coefficient of agreement for an individual $i$ (denoted by $P_c^{(i)}$) as the probability that an examinee $i$ is assigned to the same mastery state on parallel tests $x$ and $x'$. For the case of two mastery states, this probability is given by

$$P_c^{(i)} = P(x_i \geq c, x'_i \geq c) + P(x_i < c, x'_i < c) \quad (1)$$

Where $c$ is the cut off score, $x_i$ and $x'_i$ are scores for student $i$ on the two parallel tests.

The terms on the right side of the above equation are the probability of consistent mastery/mastery and non-mastery/non-mastery decisions respectively.

The coefficient of agreement for a group of $N$ students has
been defined by Subkoviak (1976) as the mean of the individual coefficients.

\[
PC = \frac{\sum_{i=1}^{N} p_c(i)}{N}
\]  

(2)

The above equation gives the sum of the probabilities of making a consistent decision for each student \(i\) weighted by his or her probability of occurrence in the group, and so represents the expected probability of a consistent decision across group. This equation provides a group estimate, \(P_c\), of assignment consistency which is a very good measure of the reliability of a criterion referenced test and for classifying students into masters and non-masters.

In order to estimate \(P_c^{(i)}\) Subkoviak assumed that for each examinee, scores on the two forms of the criterion referenced test were independently distributed. Also he assumed that the distributions of \(X_i\) and \(X'_i\) for a fixed student are identically binomial in form. (Conditions that make this assumption reasonable are: (a) each of the \(n\) items on a test is scored 0 or 1, (b) the outcome on one item does not affect the outcome on another, and (c) the probability of a correct response remains constant across items.) With both these assumptions Equation 1 simplifies to

\[
P_{c}^{(i)} = [p(x_i > c)]^2 + [1-p(x_i > c)]^2
\]  

(3)

Where

\[
p(x_i > c) = \sum_{x_i = c}^{n} \binom{n}{x_i} p_1^{x_i} (1-p_1)^{n-x_i}
\]  

(4)
The quantity $p^*_i$ in Equation 4 is the true probability of a correct item response for student $i$, which can be estimated from his or her observed score $x_i$ on a single test consisting of $n$ items.

$$\hat{p}_i = \frac{x_i}{n}.$$  

Thus, the probability of consistent classification for each student can be estimated by using Equation 3 and 4 and for an entire group by using equation 2, with data from a single test administration.

### 2.6.9 VALIDITY OF CRITERION REFERENCED TESTS

As observed by Hambleton et al. (1978, p.38) until recently it was thought that content validity considerations of a criterion referenced test were sufficient. Cronbach (1971) and Messick (1975) have argued that to validate interpretations of criterion referenced test scores it is necessary to proceed beyond a consideration of the content validity of a test.

Sometimes a test developer will claim that a test has content validity just because the test is criterion referenced or just because it is based on a well-defined domain. Specifying the domain definition is only the initial step. A sufficient and representative number of items need to be on the test in order to claim content validity. An obvious procedure in establishing content validity involves judgements made of test items by content specialists who look for correspondence between what they judge each item to measure and the domain it purports to measure. For evaluating item-domain congruence rating scales and matching tasks can be employed.
Hambleton et al. (1978, p.39) have stated that in addition to the steps required to ensure content validity it is important to conduct construct validation studies for the intended interpretations of a set of scores. These studies will include experimental methods of investigation as well as other methods. Unfortunately, relatively little attention has been paid to the problem of construct validity of criterion referenced test scores. In one study Hambleton (1977) discussed the possibilities of Guttman scaling as well as factor analysis for studying the construct validity of criterion referenced test scores.

In some instances criterion-related validity of the criterion referenced tests have to be investigated. One example of the need for this type of validity evidence is in the context of using criterion referenced tests as post-tests for instructional units arranged in a learning sequence leading to some desired outcome (Nitko, 1983, p.459-460). A measure of the final learning outcome to be attained serves as the criterion, and criterion-related validity studies are conducted to validate using each unit's criterion referenced post-test as a mastery test that predicts attainment of the final learning outcome.

The criterion-related validity studies may be concurrent or predictive depending upon how success is defined, for example, success in the current programme, or success in the next unit of instruction.
2.7.0 USE OF ENVIRONMENT ASSESSMENTS IN CURRICULUM EVALUATION

2.7.1 BACKGROUND - NEED FOR ASSESSING EDUCATIONAL ENVIRONMENTS

Walberg and Moos (1980) have criticised teachers, educational administrators and evaluators for relying heavily, often exclusively, on tests of achievement for evaluation of students and programmes. Assessment of academic achievement cannot give a complete picture of the educational process. Hence they afford only a little insight into the causes of changes and differences in outcomes.

During the last two decades researchers like Herbert Walberg and Rudolf Moos (USA), Gary Anderson (Canada) and Barry Fraser (Australia) have done pioneering work in investigating educational environments. Their work has resulted in practical, inexpensive instruments and techniques for measuring environmental variables that account for considerable amount of the variance in learning outcomes.

Schwab (1973) has delineated the following four 'common places' which must be considered during deliberation about educational practice:

i) the learner,

ii) the teacher,

iii) the milieu or educational environment, and

iv) the subject matter.

Further, Schwab criticises curriculum efforts which have focussed predominantly on only one of these common places at the neglect of the others. By identifying educational environment as one of the four essential considerations, Schwab's work
highlights the need to incorporate the educational milieu into deliberations about educational practice.

Fraser (1981, p.70) has highlighted the need for supplementing achievement measures with classroom environment variables while in choosing criteria of curricular effectiveness.

2.7.2 THE CONCEPT OF EDUCATIONAL ENVIRONMENT

Dressel (1976, p.166) defines the term 'environment' as the conditions, circumstances and influences which surround and affect the development of an organism or a group of organisms. Another term 'climate' frequently used interchangeably with 'environment', is similarly defined as prevailing conditions affecting life or activity. According to Page and Thomas (1977, p.67) the term 'Classroom climate' refers to the 'authority pattern and social and emotional relationships within a teaching group'.

Both these definitions, differentiate organisms and life from conditions, circumstances or influences. This differentiation corresponds to a concept of environments and people interacting in or through processes designed to achieve certain objectives. The environment thus includes such psychological, social and physical components as those listed by Dressel. (1976, p.166)

* Campus mores, traditions, rules
* Acceptable standards of behaviour and achievement
* Innovative – Conservative balance
* Grounds, architecture and facilities
* Value orientation and priorities
People and an environment are essential to create processes (through a planned schedule of interactions among people in the environment) to produce desired outcomes.

Pace (1969) states that an environment is what people perceive it to be.

Different terms have been used by different researchers to refer to the environments of educational settings. For example researchers like Anderson (1973) and Fraser (1981) have used the term 'Learning Environment'. Moos and Trickett (1974) have used the term 'Classroom Environment'. Barclay (1971) has employed the term 'Classroom Climate'. The term 'Educational Environment' has been used by Walberg (1979). The four terms cited above have been used to refer to the psychosocial dimensions of the environments in educational settings.

Wittrock (1970, p.7) has used the term 'instructional environment' which emphasizes the instructional aspects of educational settings. He has stated that if one wishes to evaluate a given course he has to look at the instructional environment provided in the course.

2.7.3 DISTINCTION BETWEEN SCHOOL (INSTITUTION) LEVEL AND CLASSROOM LEVEL ENVIRONMENT:

Many Researchers (Anderson, 1982; Fraser and Rentoul, 1982; Genn, 1984) have drawn a distinction between classroom-level environment from school-level environment.
The term 'school-environment' is more global than classroom environment. It pertains to the various aspects of the climate prevailing in a whole educational institution. Pace (1969) has used the term 'College/University environment' to refer to the institution-level environment of higher education institutions. Peterson, Centra, Hartnett and Linn (1970) have also used the term 'Institution' instead of 'School' in their study.

School-level environment research has relied heavily on earlier work on organizational climate in business contexts (Anderson, 1982). Hence it is usually associated with the field of educational administration. (Thomas, 1976).

Sharma (1971) studied organisational climate of schools in India and its relationship with pupils' achievement and reported that there was significant positive relationship between organisational climate and school academic Index. In another prediction study prognosticating school climate, Sharma (1972) found that headmaster's effectiveness is a significant predictor of organisational climate.

Fraser (1985, p.5) while reviewing two decades of research on perceptions of classroom environment has observed as follows:

"Whereas classroom-level research has been concentrated on secondary and primary schools rather than in higher education, a sizable proportion of school-level environment research has involved the climate of higher education institutions".
2.7.4 A MODEL OF THE ROLES OF LEARNING ENVIRONMENT IN CURRICULUM EVALUATION

The importance of the learning environment in curriculum evaluation is implicitly endorsed in several popular evaluation models. For example, Stake's (1967) countenance model emphasizes the need to consider "transactions", the Context-input-process-product (CIPP) model (Stufflebeam et al. 1971) highlights the importance of "processes", and Parlett and Hamilton's (1976) illuminative model stresses the place of the "milieu" in evaluation.

Fraser (1981, p.4) has developed a heuristic model to represent the vital role of environment variables in curriculum evaluation. This model involves the following four distinct construct domains:

- i) **the curriculum** to be evaluated
- ii) **the aptitude** or individual differences of the students following the curriculum
- iii) **the learning environment** of classrooms in which the curriculum is used
- iv) **student learning outcomes**.

The model presented in Figure 2.4 highlights the important role environmental variables play both as dependent and independent variables in curriculum evaluation research.
Fraser (1981, p.6) advocates the use of classroom environment variables in the role of both dependent and independent variables in the same curriculum evaluation study. This suggestion is consistent with Goodlad’s (1972) paradigms and with Anderson and Walberg’s (1974) claim that differences among curricula are often reflected first and most strongly in changes in learning environment and that, later and in moderated form, these changes show up in terms of student learning outcomes.

When viewed as independent variables, classroom environment dimensions constitute one of three major classes of determinants of student learning (Curriculum and aptitude are the other two).

Lewin's (1936) work on field theory, as Fraser (1981) points out, forms the basis and provides support for the use of environment dimensions as predictors of learning outcomes. Lewin recognised that both the environment and its interaction with
personal characteristics of the individual are potent determinants of human behaviour. This is expressed in his formula given below:

\[ B = f(P, E). \]

This formula stresses the need for new research strategies in which behaviour (B) is considered a function (f) of the person (P) and the environment (E).

When employed as dependent variables, environment dimensions provide process criteria of curricular effectiveness. (Walberg 1975; Eggleston and Galton, 1976). The use of these process criteria is very important because it is becoming common for curriculum developers to define, not only the objectives to be achieved by students, but also the nature of the learning environment considered desirable.

Although the model shown in Figure 2.4 depicts the direct influence on learning of the three domains of curriculum, aptitude and environment, interactions among domains also have to be considered in curriculum evaluation. The three possible interactions have been briefly described below:

2.7.41 Interaction between curriculum and aptitudinal variables:

These interactions have been focussed in a substantial number of investigations referred to as ATI (aptitude - treatment - interaction) studies (e.g. Salomon, 1972; Cronbach and Snow, 1977). The purpose of ATI studies has been to explore the differential effectiveness of alternative educational treatments for students differing on some aptitudinal variables.
2.7.42 Curriculum-environment interactions:

Fraser (1981, p. 43) has stated that over the last decade relatively few curriculum environment interaction studies have been undertaken and these have failed to provide strong evidence supporting the fruitfulness of this line of research. Walberg's (1970) work with Hardward project Physics and Fraser's (1979) study of the Australian Science Education Project have revealed the existence of some curriculum environment interaction which accounted for some substantial increments in outcome variance.

2.7.43 Interactions between aptitude and environment:

Past research in this area has concentrated on investigation of associations between student outcomes and the nature of the actual environment.

Recently number of studies (e.g. Rich and Bush, 1978; Fraser and Rentoul, 1980 and Fraser and Fisher, 1983b) have explored whether students achieve better when there is a higher similarity between the actual classroom environment and that preferred by students. Such research studies are referred to in the literature as Person-environment fit research (Hunt, 1975). The findings of the studies in this area suggest that actual preferred congruence is more important than the actual environment per se in predicting learning outcomes.

The model in Figure 2.4 highlights the need for incorporating classroom environment dimensions as both predictor and criterion variables in curriculum evaluation. Fraser (1981, p. 6) has stated that apart from handful of notable exceptions
(eg. Cort, 1979; Fraser, 1979), the use of environment variables in the role of both independent and dependent variables in the same curriculum evaluation study has been almost non-existent. Thus Fraser makes a compelling case for the vital importance of learning environments in curriculum evaluation.

2.7.5 VARIOUS APPROACHES TO STUDYING EDUCATIONAL ENVIRONMENTS

Three distinct methodologies for assessing and studying classroom environments have been described by Fraser and Walberg (1981, p.67). They are:

i) Interaction analysis

ii) Naturalistic inquiry and case study, and

iii) Use of student and teacher perceptions of the classroom.

Interaction analysis involves observation and systematic coding of classroom communication according to some category system. (eg. Flanders, 1970; Dunkin and Biddle, 1974; Peterson and Walberg, 1979). Power (1977) has reviewed a number of interaction studies carried out in Science Classrooms.

The second approach to studying classroom environment involves application of the techniques of naturalistic inquiry, ethnography and case study. Relatively few studies of classroom environments have used the technique of naturalistic inquiry. 'Case Studies in Science Education' (Stake and Easley, 1978) provides one of the few instances of the use of naturalistic inquiry methods in Science classrooms.

The third approach to studying classroom environment involves the use of students' and teachers' perceptions of the
various characteristics of the classroom. It is this third approach which forms the basis of assessment of learning environment in the present study.

In this section the method of assessing classroom environment in terms of students' and teachers' perceptions is compared with alternative approaches and the relative merits of perceptual measures are weighed.

2.7.51 Student Perceptual Measures vs Observational Techniques

Jessor and Jessor (1973) distinguishes between the "objective" approach of directly observing the environment (as is done in interaction analysis) and the "subjective" approach based on milieu inhabitants' apprehension of the environment. A similar distinction has been made by Murray (1938). He used the term alpha press to describe the environment as assessed by a detached observer and the term beta press to describe the perceived environment of milieu inhabitants.

Borich and Madden (1977, p.7) have suggested the use of 'Level of inference' as a dimension to guide the search for methods and tools for assessment of environments. Their suggestion is based on Rosenshine's (1970) distinction between low inference and high inference measures of classroom environment.

High-inference measures pertain to the general characteristics of the classroom events (eg. Order and Organisation of the class) or teachers (eg. Teachers' resourcefulness) or students (eg. Student preparedness). High inference measures require the respondent generalise the events, and then judge. Low inference
measures pertain to discrete and easily observable units of activity such as the number of questions asked by the teacher or the number of times the teachers reinforced the students. Low inference measures are so clearly defined that little inference is required of the observer.

Classroom observation schemes focus upon low inference variables; but perceptual measures focus on high inference variables.

2.7.52 Advantages of Student Perceptual Measures over Observational techniques:

Fraser and Walberg (1981, p.67-68) have identified the advantages which student perceptual measures have over observational techniques. They are listed below:

(i) Paper-and-pencil perceptual measures are more economical than classroom observation techniques which involve the expense of trained outside observers. (As an extreme example for the high cost of classroom observational technique, the evaluation and improvement of the Physics teaching materials produced in the U.S.A. by the Physical Science Study Committee (PSSC) is reported by Marsh (1964) to have cost about three times as much as the original production).

(ii) Perceptual measures are based on students' experiences over many lessons, while observational data usually are restricted to a very small number of lessons.

(iii) Perceptual measures involve the pooled judgements of all students in a class whereas observation techniques typically involve only a single observer.
Students' perceptions, because they are the determinants of student behaviour more so than the real situation, can be more important than observed behaviours. This is consistent with Thomas's (1928, p572) apothegm: If men define situations as real, they are real in their consequences.

Perceptual measures of classroom environment typically have been found to account for considerably more variance in student learning outcomes than have interaction variables. (eg. Rosenshine, 1969; Fiedler, 1975).

Further, students' ability to perceive and weigh classroom stimuli in order to make valid judgements about psychosocial characteristics of their classroom is found to be quite good by Walberg and Haertel (1980). They also suggest that the molar judgements made by students could mediate the multiplicity of molecular events of instruction and activities in the classroom.

Choosing an Appropriate Level of Analysis:

The two commonly used units of statistical analysis in prior classroom environment research have been the individual and the class mean.

Stern, Stein and Bloom (1956, p.37) have extended Murray's classification of alpha and beta press by subdividing beta press into two types viz., private beta press and consensual beta press. The idiosyncratic view that each person has of the environment is called private beta press. The shared view that members of a group hold about the environment is called consensual beta press.
Fraser (1985, p.5) has noted that private and consensual beta press could differ from each other and both could differ from the detached view of alpha press of a trained nonparticipant observer.

While designing classroom environment studies a researcher must decide whether his analysis will involve the perception scores obtained from individual students (private beta press) or whether these will be combined to obtain the average of the environment scores of all students within the same class. (Consensual beta press).

The choice of unit of analysis is of key importance for the following four reasons (Fraser 1985, p.5):

(i) Perceptual measures having the same operational definition may have different substantive interpretations with different levels of aggregation. (e.g., the difference between private and consensual press).

(ii) Relationships obtained using one unit of analysis could differ in magnitude and even in sign from those obtained using another unit. (e.g., Robinson, 1950). Furthermore, Haertel, Walberg and Haertel's (1981) meta analysis of classroom environment studies has revealed that there was a monotonic increase in the magnitude of correlations with increasing aggregation from student to subgroup to class to school.

(iii) The use of certain units of analysis (e.g., individuals when classes are the primary sampling units) violates
the requirement of independence of observations. This violation affects the results of any statistical significance tests because an unjustifiably small estimate of the sampling error is used. (Peckham, Glass and Hopkins, 1969; Ross, 1978). To solve this problem researchers like Keeves and Lewis (1983) have used the individual as the unit of analysis but employed the Jack-Knife technique (Mosteller and Tukey, 1977) to adjust significance levels to allow for nonindependence of observations.

(iv) The use of different units of analysis involves the testing of conceptually different hypotheses. Fraser's (1985, p.5) illustration of this point is very useful to researchers. According to him a study of the effects of classroom environment on some student outcome measure, employing the individual as the unit of analysis (i.e. a between student analysis) will be useful to determine the relationship between individual students' outcomes and their environment scores when class membership is disregarded. Employing the 'deviation of a student's score from the class mean' as the unit of analysis (i.e. a pooled within-class analysis) will be useful to find out whether the extent of deviation of a student's environment score from that of his classmates is related to the extent of deviation of his performance (learning outcome) score from the class mean. Employing the class mean as the unit of
analysis (i.e., a between class analysis) will be useful to determine whether the relationship between class means on the outcome measure varies with the average environment perceptions of the students within a class.

2.7.6 INSTRUMENTS FOR ASSESSING CLASSROOM ENVIRONMENT

Although a large number of tests are available for measuring student outcomes, very few instruments are available for assessing the perceptions of classroom environments.

The first part of this section deals with Instruments used for assessing Psychosocial aspects of classroom environment and the second part deals with instruments used for assessing instructional Aspects (like cognitive press, teaching-learning process) of classroom environment.

2.7.61 Instruments for Assessing Psychosocial Aspects of Classroom Environment:

The background and nature of five main instruments commonly used in prior research to assess the Psychosocial aspects of classroom environments is presented in this section. The instruments considered are listed below:

i) Learning Environment Inventory (LEI)

ii) Classroom Environment Scale (CES)

iii) Individualized Classroom Environment Questionnaire (ICEQ)

iv) My Class Inventory (MCI)

v) College and University Classroom Environment Inventory (CUCEI).
The scales in the instruments listed above have been carefully developed, extensively field tested and shown to be reliable.

Table 2.1 provides an overview of scales contained in the LEI, CES, ICEQ, MCI and CUCEI. This table furnishes the name of each instrument and its original author(s), the level it is suited for, the number of items contained in each scale, the total number of scales in each instrument and the classification of each scale according to Moos's (1974) scheme for classifying the psychosocial aspects of human environments. (cf. 3.5.422)

A noteworthy feature of most of the instruments in Table 2.1 is that they have two forms i.e. one to measure perceptions of actual classroom environment, and another to measure perceptions of preferred (or ideal) classroom environment.

Further, short forms of the ICEQ, MCI and CES have also been developed (Fraser, 1982a; Fraser and Fisher, 1983b) to provide more rapid and economical instruments.

2.7.62 Instruments for Assessing Instructional Aspects of Classroom Environment

As observed by Fraser (1981, p.28) psychosocial aspects of classroom environment have been measured most often with perceptual instruments, whereas teaching behaviour (or instructional aspects of classroom environment) has been measured usually with classroom interaction techniques.
<table>
<thead>
<tr>
<th>Instrument and its original author(s)</th>
<th>Level</th>
<th>Items</th>
<th>Total No. of Scales</th>
<th>Relationship dimensions</th>
<th>Personal development dimensions</th>
<th>System maintenance and Change Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Learning Environment Inventory (LEI)</td>
<td>Secondary</td>
<td>7</td>
<td>15</td>
<td>Cohesiveness</td>
<td>Speed</td>
<td>Diversity, Formality</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Friction</td>
<td>Difficulty</td>
<td>Material Environment, Goal Direction, Disorganization, Democracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Favoritism</td>
<td>Competitive-ness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cliqueness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Satisfaction</td>
<td>Apathy</td>
<td></td>
</tr>
<tr>
<td>2. Classroom Environment Scale (CES)</td>
<td>Secondary</td>
<td>10</td>
<td>9</td>
<td>Involvement</td>
<td>Task Orientation</td>
<td>Order and Organization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Affiliation</td>
<td>Competition</td>
<td>Rule Clarity, Teacher Control, Innovation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teacher Support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Individualized Classroom Environment Questionnaire (ICEQ)</td>
<td>Secondary</td>
<td>10</td>
<td>5</td>
<td>Personalization</td>
<td>Independence</td>
<td>Differentiation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Participation</td>
<td>Investigation</td>
<td></td>
</tr>
<tr>
<td>4. My Class Inventory (MCI)</td>
<td>Elementary</td>
<td>6-9</td>
<td>5</td>
<td>Cohesiveness</td>
<td>Difficulty</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Friction</td>
<td>Competitive-ness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Satisfaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. College and University Classroom Environment Inventory (CUCEI)</td>
<td>Higher Education</td>
<td>7</td>
<td>7</td>
<td>Personalization</td>
<td>Task Orientation</td>
<td>Innovation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Involvement</td>
<td>Student-Teacher-Cohesiveness</td>
<td>Individualization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cliqueness</td>
<td>Satisfaction</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Fraser (1985, p.38)
However researchers like Eash, Rasher and Waxman (1979) and Stayrook, Corno and Wynne (1978) have departed from this tradition by developing instruments which measure student perceptions of teacher behaviours.

A survey of the literature indicates that only very few perceptual instruments are available for assessing the instructional aspects (like cognitive press and teaching-learning process which encompasses instruction, evaluation, diagnosis and guidance) of classroom environments.

The background and nature of four perceptual instruments used in prior research to assess the instructional aspects of the classroom environment are discussed, in this section. Table 2.2 furnishes the names of these four instruments, respective original author(§), the total number of scales in each and the names of these scales.

All the four instruments listed in Table 2.2 are suitable for school level institutions. They have been field tested and shown to be reliable.

Eash and Colleagues' instrument measures nine teacher behaviours which have been found to be empirically linked with student learning.

In the Class Activities Questionnaire, the first two major scales namely, Lower Thought Processes and Higher Thought Processes are based on the Bloom et al. (1956) taxonomy of objectives in the cognitive domain. As observed by Fraser (1981, p.27) these scales focus mainly on cognitive press.
<table>
<thead>
<tr>
<th>Instrument and its Original author(s)</th>
<th>No. of Scales</th>
<th>Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Perception Scale (TPS)</td>
<td>3</td>
<td>Student perception of three teacher behaviours listed below:</td>
</tr>
<tr>
<td>Stayrook, Corno and Wynne (1978)</td>
<td></td>
<td>Structuring</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Soliciting and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reacting</td>
</tr>
<tr>
<td>Instrument for evaluating teacher behaviour from student perceptions</td>
<td>9</td>
<td>Clarity; variability in teaching methods and materials;</td>
</tr>
<tr>
<td>Eash, Rasher and Waxman (1979)</td>
<td></td>
<td>Enthusiasm; Task-Oriented behaviour;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indirectness; Student opportunity to learn the material;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teacher use of structuring comments;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple levels of questions and cognitive discourse; and teacher criticism.</td>
</tr>
<tr>
<td>Class Activities Questionnaire (CAQ)</td>
<td>Four major scales</td>
<td>Lower Thought Processes</td>
</tr>
<tr>
<td>Steele, House and Kerins (1971)</td>
<td>and 16 sub-scales</td>
<td>(Subscales: Memory, Translation and Interpretation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Higher Thought Processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Subscales: Application, Analysis, Synthesis and Evaluation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Classroom Focus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Subscales: Discussion, Test-Grade Stress and Lecture)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Classroom Climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Subscales: Enthusiasm, Independence, Divergence, Humour, Teacher talk and Homework)</td>
</tr>
<tr>
<td>Instrument for measuring teachers' or external observers' perceptions of openness in Classroom Environment</td>
<td>8</td>
<td>Provision for learning; Humaneness, Respect, Openness and Warmth; Diagnosis of learning events; Instruction, Guidance and Extension of learning; Evaluation of Diagnostic Information; Seeking opportunities for Professional Growth; Self-Perception of teacher; Assumptions about children and learning process.</td>
</tr>
<tr>
<td>Walberg and Thomas (1972)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.7.7 RESEARCH USING ENVIRONMENT ASSESSMENTS IN CURRICULUM EVALUATION:

To illustrate the dual roles of environmental dimensions as predictor and criterion variables in curriculum evaluation, past studies are presented in two separate sections. The first section deals with predictive validity studies of associations between perceptions of environment and student learning outcomes. The second section reviews studies of the criterion validity of classroom environment measures in differentiating between classrooms following alternative curricula. In the context of the present study the latter approach is only of theoretical interest.

2.7.7.1 Review of Predictive Validity Studies:

When classroom environment characteristics are employed as independent variables, research results yield information about the predictive validity of environmental perceptions in accounting for variance in student learning outcomes. Evidence accrued from a large number and variety of predictive validity studies reviewed by Fraser (1981, 1985) indicates that student perceptions of their classroom environments account for appreciable amounts of variance in learning outcomes often beyond that attributable to background student characteristics such as pretest performance, general ability or both.

Previous research studies of associations between outcome measures and classroom environment perceptions span various countries, grade levels and subject areas. Furthermore, previous research reflects many methodological variations including choice
of classroom environment instrument, Choice of learning outcomes and instruments to measure them, sample size, unit of statistical analysis, background variables controlled and data analysis techniques.

Table 2.3 presents an overview of eleven research studies of predictive validity of student environment perceptions in Science Classrooms. The first eight studies illustrate the research carried out in developed countries. The last three studies illustrate the research conducted in developing countries.

2.7.711 Predictive Validity Studies in Developed Countries:

The first part of this section deals with research associated with Harvard Project Physics and the second part presents other predictive validity studies conducted in developed countries.

(A) Predictive Validity Studies related to Harvard Project Physics

As shown in Table 2.3 some of the earliest research into the predictive validity of students' classroom environment perceptions was a series of studies conducted as part of the research and evaluation activities of Harvard Project Physics. This was an experimental Physics course for Grade 10 to 12 students. This series of studies is important because it illustrates the use of a variety of important methodological techniques. As observed by Fraser (1981, p.33) these studies illustrate the use of a variety of interaction and curvilinear environment terms and various multivariate data analysis techniques. These studies followed a system of data collection
### TABLE 2.3
Overview of Studies of Predictive Validity of Student Environment Perceptions in Science Classrooms

<table>
<thead>
<tr>
<th>SNo.</th>
<th>Study</th>
<th>Instruments and Measures</th>
<th>Sample</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Environment</td>
<td>Learning outcomes</td>
<td>Location</td>
</tr>
<tr>
<td>1.</td>
<td>Studies related to Harvard project physics: Walberg and Anderson (1968); Anderson and Walberg (1968); Anderson (1970); Walberg (1969 a, b, 1972)</td>
<td>Learning Environment Inventory (LEI) (final or preliminary versions)</td>
<td>Selected from: achievement; understanding of nature of science; science processes; participation in physics activities; science interest; attitudes</td>
<td>Mainly national U.S. sample; also Canadian classes in some studies</td>
</tr>
<tr>
<td>2.</td>
<td>Lawrenz (1976)</td>
<td>LEI</td>
<td>science attitudes</td>
<td>Midwest regions of U.S.A.</td>
</tr>
<tr>
<td>3.</td>
<td>Fraser (1978, 1979)</td>
<td>LEI (modified)</td>
<td>3 inquiry skills; understanding of nature of science; several attitudes</td>
<td>Melbourne, Australia</td>
</tr>
<tr>
<td>Sl. No.</td>
<td>Study</td>
<td>Instruments and Measures</td>
<td>Sample</td>
<td>Methodology</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------------------------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Environment</td>
<td>Learning outcomes</td>
<td>Location</td>
</tr>
<tr>
<td>4.</td>
<td>Power and Fisher (1979)</td>
<td>LEI (modified) and Class Activities Questionnaire</td>
<td>2 achievement; 3 attitude; 3 satisfaction</td>
<td>Melbourne, Australia</td>
</tr>
<tr>
<td>5.</td>
<td>Fraser and Fisher (1982b)</td>
<td>My Class Inventory (MCI)</td>
<td>inquiry skills; understanding of nature of science; attitude</td>
<td>Tasmania, Australia</td>
</tr>
<tr>
<td>6.</td>
<td>Rentoul and Fraser (1980)</td>
<td>Individualized Classroom Environment Questionnaire (ICEQ)</td>
<td>2 inquiry skills, enjoyment</td>
<td>New South Wales, Australia</td>
</tr>
<tr>
<td>8.</td>
<td>Fraser and Fisher (1982a)</td>
<td>Classroom Environment Scale (CES) and ICEQ</td>
<td>3 cognitive and 6 affective measures</td>
<td>Tasmania, Australia</td>
</tr>
<tr>
<td>Sl. No.</td>
<td>Study</td>
<td>Environment</td>
<td>Learning outcomes</td>
<td>Location</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------------</td>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>9.</td>
<td>Walberg, Singh and Rasher (1977)</td>
<td>LEI (Hindi version)</td>
<td>achievement</td>
<td>Rajasthan</td>
</tr>
<tr>
<td>10.</td>
<td>Chatlyanonda (1978)</td>
<td>LEI (Thai version)</td>
<td>3 attitudes</td>
<td>Near Bangkok</td>
</tr>
<tr>
<td>11.</td>
<td>Hofstein et al (1979)</td>
<td>LEI</td>
<td>4 attitudes</td>
<td>Israel</td>
</tr>
</tbody>
</table>

Adapted from Fraser and Walberg (1981)
(Walberg and Welch, 1967) in which testing time was reduced by having different random subsamples of students within a given class respond simultaneously to different parts of a total test battery. The design of this series of studies in general, involved the prediction of learning outcomes measured at the end of a school year from student perception of the classroom environment measured during the year, usually with statistical control for performance on corresponding beginning-of-year learning outcome measures.

The predictive validity of the LEI was supported in this series of studies using three different combinations of units of analysis; these were individual student environment perceptions and individual learning outcome scores (Walberg and Anderson, 1968), class mean environment perceptions and class mean learning outcome scores (Anderson and Walberg, 1968) and class mean environment perceptions and individual learning outcome scores (Anderson, 1970).

B. Other Predictive Validity Studies in Developed Countries

This section presents a review of seven selected predictive validity studies conducted in science classrooms in developed countries. (cf. Item No.2 to 8 of Table 2.3).

Lawrenz's (1976) study of 238 classes in the United States of America showed that the amount of variance in raw attitude scores accounted for by a block of 10 LEI scales was 39 percent for biology classes, 32 percent for chemistry classes and 29 percent for physics classes.
In Australia, Fraser (1978) using a modified nine scale version of the LEI found that more favourable classroom environments tended to promote more positive attitudes to experiments as a source of scientific information. Less favourable environments tended to promote more positive attitudes to three nonexperimental and more authoritarian sources, viz., experts, books and teachers.

Power and Tisher (1979) conducted another study involving Junior High School Science classes in Australia. They used a modified version of the LEI together with the Class Activities Questionnaire. This study revealed the existence of a number of statistically significant relationships between learning outcomes and environment dimensions.

Using the MCI - the simplified version of the LEI - Fraser and Fisher (1982b) conducted a study involving twelve year old students of 100 science classes at Tasmania in Australia. Results from the different types of analyses conducted by them consistently supported the predictive validity of MCI.

Rentoul and Fraser (1980) have reported the first use of the ICEQ (student actual form) in a predictive validity study involving a sample of science and social science classes. This study revealed that student perceptions on the five ICEQ scales together accounted for a significant increment in the variance in an affective outcome, but not in two cognitive outcomes (beyond that attributable to background student characteristics).

Gardner (1974, 1976) used a modified version of eight environment scales selected from Stern's (1970) College
Characteristics Index in predicting four attitudinal criteria from students' classroom perceptions. Analysis of covariance techniques with the student as the analytic unit revealed that the number of significant associations between an attitudinal outcome and a classroom environment dimension was five times that expected by chance. One of the interesting results that emerged out of this study carried out in Melbourne, Australia is that students expressed greater enjoyment of Physics learning in classrooms perceived as highly achievement oriented, cognitively well-organized, intellectual, stimulating and physically well-organized.

Fraser and Fisher (1982a) have reported the first use of the Classroom Environment Scale (CES) specifically in science classrooms. They investigated the relationships between students' affective and cognitive outcomes and their perceptions of classroom psychosocial environment as measured by the CES and the ICEQ. This study used the data base from the sample of science classes at Tasmania in Australia. The data collected were analyzed in six different ways namely simple, multiple, and canonical correlation analysis performed separately for raw post-test scores and residual post-test scores adjusted for corresponding pretest and general ability. The results obtained revealed sizable environment-outcome associations. Further analyses showed that the CES and ICEQ made appreciable, unique contributions to explaining outcome variance and that the magnitudes of environment-outcome relationships were larger when the class was employed as the unit of analysis than when the student was used.
2.7.712 Predictive Validity Studies in Developing Countries

Three studies (cf. Item No.9 to 11 of Table 2.3) carried out in developing countries are discussed in this section. Translated versions of the LEI have been employed in predictive validity studies involving senior high school science classes in India (Walberg, Singh and Rasher, 1977) and Thailand (Chatiyanonda, 1978) and Israel (Hofstein, Gluzman, Ben-Zvi and Samuel, 1979). All the three studies provided convincing and consistent evidence to support the cross-cultural predictive validity of students' classroom environment perceptions.

Hofstein, Gluzman, Ben-Zvi and Samuel (1979) used a modified version of 13 of the LEI scales in investigating attitude towards Chemistry among 400 eleventh grade students from 12 classes in eight Israeli Schools. Attitude to Chemistry was measured by four factor-analytic scales (attractive and exciting, clear and understandable, necessary and useful, and inexact and confusing). A canonical analysis using the student as the unit of analysis was performed. Findings of this study indicated that attitudes to Chemistry were more favourable in classes perceived as having higher goal direction and satisfaction and lower disorganisation, difficulty, friction and speed.

Predictive Validity Study in India

In India, Walberg, Singh and Rasher (1977) have conducted a cross cultural replication study which involved administration of the Learning Environment Inventory translated into Hindi. The study was carried out in the State of Rajasthan involving a
random sample of 3000 students of tenth grade in 83 Science and 67 Social Science Classes. The mean score of the subgroup within the class formed by grouping separately studious class members was used as the unit of statistical analysis. The magnitude of the simple correlation between raw end-of-course achievement scores and different LEI scales ranged from 0.41 to 0.70 for Science and from 0.58 to 0.81 for Social studies. When IQ was partialled out of the relationship, these correlations ranged from 0.17 to 0.57 for Science and from 0.36 to 0.73 for Social studies. Results of multiple regression analyses revealed that student perceptions on the block of 15 LEI scales accounted for a significant increment of 28 percent of science achievement variance and 44 percent of social studies achievement variance over and above that attributable to general ability.

2.7.713 Meta Analysis of Predictive Validity Studies

The general pattern of findings in the predictive validity studies is reflected in the set of studies reviewed by Anderson and Walberg (1974a). Their analysis reveals that student perception of classroom environment accounted for between 13 and 46 percent (median of 30 percent) of the variance in cognitive, affective and behavioural post-course measures beyond that accounted for by parallel precourse measures.

The findings from previous research on predictive validity are also summarised in a meta analysis by Haertel, Walberg and Haertel (1981). Their meta-analysis of outcome-environment studies involved 634 correlations from a collection of 12 studies.
of 10 data sets from 823 classes in eight subject areas containing 17,805 students in four nations. Their findings are:

(1) Both learning post-test scores and regression adjusted gains in learning in a variety of subject areas were positively associated with student-perceived cohesiveness, Satisfaction, Task Difficulty, Formality, Goal Direction, Democracy and the Material Environment, and negatively associated with Friction, Cliqueness, Apathy and Disorganisation.

(ii) The magnitude of the outcome-environment correlation was significantly associated with the dimension of classroom environment considered, the unit of statistical analysis and the nation in which the study was conducted. However it did not depend upon the number of students tested, the subject matter taught, the type of learning outcome, or the presence or absence of statistical control for pretest or general ability.

(iii) Correlations were generally higher in samples of older students and in studies using collectivities such as classes and schools (in contrast to individual students) as the units of statistical analysis.

The evidence obtained from the predictive validity studies reviewed confirms that classroom environment is a major determinant of learning outcomes. Therefore, it is desirable to include classroom environment dimensions as independent variables in future studies which attempt to evaluate effectiveness of curricula in terms of their impact on student learning.

2.7.72 Review of Criterion Validity Studies

Classroom environment dimensions have been used as dependent variables in the past research for the following three purposes:

(i) To evaluate curricula

(ii) To study the differences between student and teacher perceptions of actual and preferred environment, and
(iii) To investigate the relationships between classroom environment variables and other factors like Type of School, Class size or Grade level.

As noted earlier, in the context of analytical concern of the present study, the criterion validity studies which seek to judge whether curricula cause specific changes in the learning/instructional/classroom environment, are only of theoretical interest. Owing to their very limited relevance to the present investigation, only a brief review has been presented in this section.

Curriculum evaluation studies cited here have used perceptual measures of classroom environment dimensions as criterion or dependent variables. When employed as dependent variables, classroom environment dimensions provide important process criteria of curricular effectiveness (Walberg, 1975; Eggleston and Galton, 1976; Fraser 1981). As observed by Fraser (1981, p.47) compared to the larger number of predictive validity studies, there is a relative shortage of studies using classroom environment as a criterion of curricular effectiveness.

Table 2.4 presents an overview of a few curriculum evaluation studies using classroom environment perceptions as criterion variables.

Using the penultimate version of Harvard Project Physics (HPP) materials from classes using alternative Physics curriculum materials, Anderson, Walberg and Welch (1969) studied students' perceptions on the LEI in 150 senior high school classes. A multiple discriminant analysis with the class mean as the unit of
### TABLE 2.4
Overview of Curriculum Evaluation Studies Using Classroom Environment Perceptions as Criterion Variables

<table>
<thead>
<tr>
<th>Study</th>
<th>Instrument</th>
<th>Independent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraser (1976, 1979); Tisher and Power (1976, 1978); Power and Tisher (1979);</td>
<td>LEI</td>
<td>Use of Australian Science Project.</td>
</tr>
<tr>
<td></td>
<td>CAQ</td>
<td></td>
</tr>
<tr>
<td>Eash, Rasher and Walberg (1977)</td>
<td>LEI</td>
<td>Use of innovative middle school curriculum</td>
</tr>
<tr>
<td>Levin (1980)</td>
<td>LEI</td>
<td>Use of individualized curriculum</td>
</tr>
<tr>
<td>Cort (1979)</td>
<td>LEI</td>
<td>Use of MACOS (Man: A Course of Study) vs. alternative materials</td>
</tr>
<tr>
<td></td>
<td>CAQ</td>
<td></td>
</tr>
<tr>
<td>Baba and Fraser (1983)</td>
<td>Locally developed</td>
<td>Use of new Fijian social science curriculum</td>
</tr>
<tr>
<td>Kuhlemeyer (1983); Wierstra (1984)</td>
<td>ICEQ</td>
<td>Use of new Dutch physics curriculum</td>
</tr>
</tbody>
</table>

Adapted from Fraser (1985).
analysis revealed that students in classes using Harvard Project Physics materials perceived their classes as more diverse and democratic, less difficult and goal directed, and having a better physical environment and less friction.

Fraser (1976, 1979) in his evaluation of materials developed by the Australian Science Education Project (ASEP), employed a modified version of the LEI with a sample of 541 seventh grade students in Melbourne, to compare the perceived environment in ASEP and conventional classrooms. When student socioeconomic status, general ability and sex were controlled, multiple regression analysis revealed that ASEP students perceived their classrooms as more satisfying, more individualized and having a better material environment. In another study involving the evaluation of the ASEP, Tisher and Power (1976, 1978) and Power and Tisher (1979) followed changes occurring in student perceptions on the LEI and the Class Activities Questionnaire during the use of an ASEP unit in 20 junior high school classrooms.

Classroom environment dimensions have been employed as criterion variables in evaluating an innovative middle school curriculum in Illinois. (Eash, Rasher and Walberg, 1977). Studying 600 students allocated randomly to an innovative school and three other junior high schools, they found that students attending the innovative school had significantly more favourable perceptions than control group students on 12 of the LEI's 15 scales.
Levin (1980), evaluating an individualized curriculum in primary schools in Israel, studied 57 classes of which, 43 used an individualised instructional strategy while 14 control classes followed a traditional instructional strategy. Students in individualised classes perceived significantly greater autonomy than students in traditional classes.

Cort (1979) studying the differences in the students' perception of classroom environment in classes following MACOS (Man: A course of study) materials found significant differences between MACOS and non-MACOS classes.

Baba and Fraser (1983), Kuhlmeier (1983) and Wierstra (1984) have likewise employed student perceptions of classroom environment as criterion variables in studying curricular effectiveness.

The criterion validity studies reviewed in this section confirm that classroom environment dimensions, when used as dependent variables, are able to differentiate revealingly, and appreciably between classrooms following alternative curricula. As noted by Fraser (1981, p.53) though the number of previous studies using classroom environment variables as criteria of curricular effectiveness is small, the evidence accrued from these studies emphasises the desirability of using classroom environment variables as criteria in curriculum evaluation.

2.7.73 Research into Differences in Perceptions on Various Forms of the same Classroom Environment Instrument

The fact that some classroom environment instruments (e.g., ICEQ) have four different forms (i.e., student actual,
teacher actual, student preferred and teacher preferred) permits investigation of differences between students and teachers in their perceptions of the same actual classroom environment and of differences between the actual environment and that preferred by students or teachers.

A research into differences between forms has been reported by Fisher and Fraser (1983). Their study involved the use of four forms of the ICEQ with a sample of 116 classes of students and 56 teachers of these classes in Tasmania. The findings of this study are:

i) Students preferred a more positive classroom environment than was actually present for all five ICEQ dimensions.

ii) Teachers perceived a more positive classroom environment than did their students in the same classrooms on four of the five ICEQ's dimensions.

As noted by Fraser (1985, p.15), these results replicate patterns emerging in other studies in school classrooms in the U.S.A. (Moos, 1979) and Australia (Fraser, 1982b).

These studies inform evaluators that students and teachers are likely to differ in the way they perceive the actual environment of the same classroom, and that the environment preferred by students commonly differs from that actually present in their classrooms.

2.7.74 Use of Environment Assessments in Formative Curriculum Evaluation

Information about environment perceptions in classes following a specific curriculum could guide attempts to change and improve the curriculum. As observed by Steele, House and
Kerins (1971) environment assessments are potentially useful in formative curriculum evaluation.

Although much research has been conducted employing classroom environment variables in summative curriculum evaluation, the literature is almost devoid of studies which have employed classroom environment instruments in formative curriculum evaluation. Fraser (1981, Ch.5), has made an original contribution to this area by illustrating the various ways in which classroom environment data can be processed to form profiles useful in guiding systematic attempts to improve classroom environments.

By assessing students' perceptions of their actual and preferred classroom environment, the discrepancies between the actual and preferred environments can be identified. Based on this, appropriate strategies to facilitate environment changes which reduce the existing discrepancies can be implemented. Profiles of students' perceptions of actual and preferred environment scores have been employed successfully in facilitating changes in the environments of colleges (DeYoung, 1977), secondary schools (Fraser, Seddon and Eagleson, 1982) and elementary schools (Fraser and Deer, 1983).

2.7.8 THRUST IN THE PRESENT STUDY

The classroom environment information provides a valuable basis for reflection upon and subsequent improvement of curriculum and teaching-learning process.

As the title of the present study suggests the main concern is providing an analytical exposition of effectiveness, in terms
of objectives, relationships and factors. In this context most of the variables including the environment variables discussed so far would obviously take the position of independent variables which would enable better description rather than prediction. The concern however is not as in the case of criterion validity studies to differentiate among the environments.

2.8.0 HYPOTHESIZED FACTORS AND SUPPORTING RESEARCH STUDIES

The research studies leading to the selection of factors hypothesized to influence the effectiveness of the Chemistry curriculum of polytechnics are presented in this section.

Based on the review of related literature a total of nineteen factors are expected to influence the effectiveness of the Chemistry curriculum under consideration.
Table 2.5 Hypothesized Factors and Research Studies Supporting their Selection

<table>
<thead>
<tr>
<th>Hypothesized Factor</th>
<th>Supporting Research Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(A) FACTORS RELATED TO THE STUDENT</strong></td>
<td></td>
</tr>
<tr>
<td>1. Chemistry Aptitude (Aptitude for the content of the curricular subject)</td>
<td>(i) Gupta (1974)</td>
</tr>
<tr>
<td></td>
<td>(ii) Stork (1981)</td>
</tr>
<tr>
<td></td>
<td>(iii) Tewari (1975)</td>
</tr>
<tr>
<td></td>
<td>(iv) Thakur (1972)</td>
</tr>
<tr>
<td>2. Proficiency in Engineering subjects</td>
<td>(i) Fletcher (1978)</td>
</tr>
<tr>
<td></td>
<td>(ii) Stork (1981)</td>
</tr>
<tr>
<td>3. Sex of the student</td>
<td>(i) Brahadeeswaran (1981)</td>
</tr>
<tr>
<td></td>
<td>(ii) Stephenson (1979)</td>
</tr>
<tr>
<td>4. Proficiency in Science in the previous course</td>
<td>(i) Brahadeeswaran (1981)</td>
</tr>
<tr>
<td></td>
<td>(ii) Steele, House &amp; Kerins (1971)</td>
</tr>
<tr>
<td></td>
<td>(iii) Taylor &amp; Colin (1979)</td>
</tr>
<tr>
<td>5. Proficiency in English language</td>
<td>(i) Barnes (1976)</td>
</tr>
<tr>
<td></td>
<td>(ii) Brahadeeswaran (1981)</td>
</tr>
<tr>
<td></td>
<td>(iii) Taylor &amp; Colin (1979)</td>
</tr>
<tr>
<td>6. Intelligence</td>
<td>(i) Howe &amp; Early (1979)</td>
</tr>
<tr>
<td></td>
<td>(ii) McBride &amp; Chiappetta (1978)</td>
</tr>
<tr>
<td></td>
<td>(iii) Stephenson (1979)</td>
</tr>
<tr>
<td>7. Student preparedness</td>
<td>(i) Brahadeeswaran (1981)</td>
</tr>
<tr>
<td></td>
<td>(ii) Steele, House &amp; Kerins (1971)</td>
</tr>
<tr>
<td></td>
<td>(iii) Taylor &amp; Colin (1979)</td>
</tr>
<tr>
<td></td>
<td>(ii) Pridmore &amp; Halyard (1980)</td>
</tr>
<tr>
<td></td>
<td>(iii) Simpson (1969)</td>
</tr>
<tr>
<td></td>
<td>(iv) Taylor &amp; Colin (1979)</td>
</tr>
<tr>
<td>9. Student involvement</td>
<td>(i) Fraser, Tregust &amp; Dennis (1984)</td>
</tr>
<tr>
<td></td>
<td>(ii) Rawat (1970)</td>
</tr>
<tr>
<td></td>
<td>(iii) Trickett &amp; Moos (1973)</td>
</tr>
<tr>
<td><strong>(B) FACTORS RELATED TO THE TEACHER</strong></td>
<td></td>
</tr>
<tr>
<td>10. Teacher preparedness</td>
<td>(i) Pohlmann &amp; Elmore (1976)</td>
</tr>
<tr>
<td></td>
<td>(ii) Simpson (1969)</td>
</tr>
<tr>
<td></td>
<td>(iii) Taylor, Reid, Holley &amp; Exon (1974b)</td>
</tr>
</tbody>
</table>
Table 2.5 (Contd..)

<table>
<thead>
<tr>
<th>Hypothesized Factor</th>
<th>Supporting Research Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Teacher's resourcefulness</td>
<td>(i) Simpson (1969)</td>
</tr>
<tr>
<td></td>
<td>(ii) Skilbeck (1976)</td>
</tr>
<tr>
<td></td>
<td>(iii) Taylor &amp; Colin (1979)</td>
</tr>
<tr>
<td></td>
<td>(iv) Wittrock (1970)</td>
</tr>
</tbody>
</table>

(C) FACTORS RELATED TO THE INSTRUCTIONAL ENVIRONMENT


These factors are classified into three groups as shown below:

A) Factors related to the student

B) Factors related to the teacher

C) Factors related to the instructional environment.

The factors included under each of the three groups are listed in Table 2.5. The names of researchers whose work supports the selection of the factor for inclusion in this study are given against each factor. The definition of each factor is presented in section 3.2.0.

2.8.1 HYPOTHESIZED FACTORS RELATED TO THE STUDENT AND SUPPORTING RESEARCH STUDIES

The review of literature led to the selection of the following nine factor related to the student, which are expected to influence the effectiveness of the chemistry curriculum of polytechnics.

1. Chemistry Aptitude

The aptitude for the content of a curricular subject, which reflects the ability - content interface of the curriculum is an important factor influencing the effectiveness of the curriculum. This fact is clearly revealed by the work of Gupta (1974) who carried out a factor analytic study of the attainments of Higher Secondary and Pre-university passed students in different aspects of Physical sciences and Mathematics. Likewise Tewari (1975) and Thakar (1972) who studied the effectiveness of specific subjects in the High School Curricula, and Stork (1981) who has designed and field-tested a model for determining the effectiveness of
Mathematics instructional programmes within the California community college system, have all found Aptitude for the content of the curricular subject as an important factor influencing the effectiveness of the curriculum. These studies led to the selection of 'Chemistry aptitude' as a hypothesized factor.

(2) Proficiency in Engineering Subjects

Fletcher (1978) has found that students who are doing well in their college academic work can be expected to do equally well in college science courses. Stork (1981, p.86) in his study pertaining to the effectiveness of Mathematics instructional programmes within the California Community College system, found that the programme for majors fully achieved the purposes of the college transfer programme, while the programme for non-major transfer students met the transfer and general education objectives, but not the goals for the area of occupational education.

In the light of the above studies and in view of the fact that Engineering subjects constitute the main course of polytechnic students and Chemistry is only a service course for them, Student's proficiency in Engineering subjects taught in the polytechnics is expected to influence their performance in chemistry.

(3) Sex of the Student

Stephenson (1979) in his investigation of the relationships between the various characteristics of students and their achievement in High school Chemistry found that boys do about as well as girls. However the researcher (Brahedeeswaran, 1981) in
his study using students of the Polytechnics in Tamil Nadu found that achievement of girls in a Diagnostic test in Chemistry was significantly higher than that of boys. This formed the basis for selecting 'Sex of the student' as a hypothesized factor.

4. **Proficiency in Science in the Previous Course**

Taylor and Colin (1979) have identified, 'Ability already possessed by the learner', as one of the factors influencing the operationalization of the intended curriculum.

The researcher (Brahadeswaran, 1981) in his study cited above (under items 3 of this section) found that achievement in chemistry of polytechnic students depended upon their proficiency in Science in their previous course. The Class Activities Questionnaire developed by Steele, House and Kerins (1971) has items to evaluate students' preparedness for the class. As polytechnic student's proficiency in science in their previous course will determine the extent to which they are prepared for the chemistry classes in the polytechnics, this has been selected as a factor influencing the effectiveness of Chemistry curriculum of polytechnics.

5. **Proficiency in English Language**

The medium of instruction in the polytechnics in Tamil Nadu is English. But in most of the High schools in Tamil Nadu the medium of instruction is Tamil. Therefore, many students joining the polytechnics will be newly exposed to instruction in English medium. Hence it is expected that students' proficiency in English language will influence their achievement in Chemistry. This rationale is also supported by the findings of a study.
conducted by the researcher (Brahadeeswaran, 1981) which revealed that achievement in chemistry of polytechnic students who had their previous education in English medium was significantly higher than that of the students who had their previous education in Tamil medium.

'Ability already possessed by the learner' which has been identified by Taylor and Colin (1979) as one of the factors influencing the effectiveness of a curriculum includes not only the prerequisite knowledge/skill in the content of the curricular subject, but also the competency of comprehending the language used as the medium of instruction.

Barnes (1976) has also emphasised the role of language in making the curriculum accessible to pupils. So, student's proficiency in English language has been included as a hypothesized factor.

6. Intelligence

Researchers like Howe and Early (1979), McBride and Chiappetta (1978) and Stephenson (1979) have stressed the significant role played by 'Students' Intelligence' in influencing their achievement of curricular objectives. Their work formed the basis for selecting 'Intelligence' as a hypothesized factor.

7. Student Preparedness

As mentioned earlier Taylor and Colin (1979) have recognised 'ability already possessed by learner' as an important factor influencing the operationalization of the intended curriculum. The class Activities Questionnaire developed by Stelle, House and
Kerins (1971) has items to evaluate learners' preparedness for the class. The findings of the study conducted by the researcher (Brahadeeswaren, 1981) reveals that the extent to which students possess the prerequisite knowledge/skills required for learning a specific task influences their performance of the task. On the basis of this rationale, 'Student preparedness' has been included as one of the hypothesized factors.

8. **Student Motivation**

Simpson (1969, p.60) has highlighted that the motivational level on which learners are operating influences instructional effectiveness. Taylor and Colin (1979) have stated that learner's eagerness to learn is an important factor that influences the efforts of teachers to achieve the objectives of the curriculum. Davis, Alexander and Yelon (1974, p.116) have identified 'Student interest in the objective/lesson' as an important variable influencing instructional effectiveness. The fact that student's motivation is one of the useful predictors of academic achievement is clearly brought out by the findings of a study conducted by Pridmore and Halyard (1980).

These observations led to the inclusion of 'Student motivation' as a hypothesized factor in this study.

9. **Student involvement**

'Involvement' has been included as an important dimension in a number of instruments like Classroom Environment Scale (Trickett and Moos, 1973) and College and University Classroom Environment Inventory (Fraser, Treagust and Dennis, 1984). These instruments have been used in a number of curriculum evaluation studies.
Rawat (1970) has used 'Student involvement' as one of the criteria for evaluating the programmes of Secondary schools in India.

Hence 'Student involvement' has been selected as a hypothesized factor.

2.8.2 HYPOTHESIZED FACTORS RELATED TO THE TEACHER AND SUPPORTING RESEARCH STUDIES

Based on the literature review the following two factors related to the teacher have been selected as factors hypothesized to influence the effectiveness of the chemistry curriculum under consideration.

1. Teacher Preparedness

Simpson (1969) has emphasised that 'Instructor preparation for class session' is an important dimension to be considered in Teacher Self-Evaluation. Pohlmann and Elmore (1976) have used Teacher preparedness as one of the dimensions in their Instructional Improvement Questionnaire. Taylor, Reid, Holley and Exon (1974b) have used 'Readiness of Staff to give time to preparing work' as an important variable in their study of constraints on the achievement of curricular aims. These observations provided the basis for including 'Teacher preparedness' as a hypothesized factor.

2. Teacher's resourcefulness

Taylor and Colin (1979) have identified a factor called 'skill and experience of the teacher' which influences the realisation of the intended aims of a curriculum. Skilbeck (1976) has highlighted the role of 'knowledge, skills and
interests of teachers' in influencing the effectiveness of a curriculum. Simpson (1969) has included a number of items assessing Teacher's resourcefulness in his list of illustrative items for Teacher Self-Evaluative instruments. Wittrock (1970, p.9) has stated that 'intellectual merit of the teachers' influences the effectiveness of instruction. These observations led to the selection of 'Teacher's resourcefulness' as a hypothesized factor.

2.8.3 HYPOTHESES FACTORS RELATED TO THE INSTRUCTIONAL ENVIRONMENT AND SUPPORTING RESEARCH STUDIES

The review of literature led to the selection of the following eight factors related to the Instructional Environment, which are expected to influence the effectiveness of the chemistry curriculum under consideration.

1. Appropriateness of Objectives

Appropriateness of curricular Objectives is reflected in two aspects viz., their usefulness or relevance and Difficulty level.

Davis, Alexander and Yelon (1974, p.116) have stated that relevance of curricular objectives plays a significant role in determining the effectiveness of the curriculum and instruction. Taylor and Maquire (1966, p.12-13) in their study showed that teachers and curriculum developers either held as valid different objectives or had different priorities within the same set of objectives. Tawney (1973, p.159-176) has emphasized the need for considering teachers' views on the relevance of curricular objectives in curriculum evaluation. Based on this rationale,
'Appropriateness of objectives' has been included as one of the hypothesized factors.

2. Adequacy of Instructional Time

Wittrock (1970, p.7) has identified 'Amount of time devoted to learning' as one of the important variables influencing the instructional environment. In a curriculum evaluation study carried out by the Technical Teachers' Training Institute (Western Region) (1982) in India, information about the number of periods actually required for teaching each topic in the curriculum has been collected from teachers and compared with the number of periods recommended in the curriculum document. This comparison has provided valuable information to modify the curriculum and make it realistic with respect to instructional time.

On the basis of the observations cited above 'Adequacy of Instructional Time' has been selected as a hypothesized factor.

3. Instructional Resources

Skilbeck (1976) has identified 'Instructional materials and resources' as one of the factors that influences the effectiveness of a curriculum. Taylor et al. (1974b) have recognised that resources and facilities in the school influence the effectiveness of teaching. Taylor and Colin (1979, p.20) have identified "Physical conditions of schools and classrooms" as an important factor that influences instructional effectiveness. Ainley (1978) has found a significant positive relationship between the standard of science facilities and student perceptions of a dimension of classroom environment.
called 'Stimulation through Variety'. Based on these studies, 'Instructional resources' has been included as a hypothesized factor.

4. Instructional Methods and Procedures

According to Zais (1976, p.354) the nature of 'instructional strategies' employed to bring the inert curriculum document to life as a functioning dynamic in the classroom setting plays a significant role in determining the effectiveness of the curriculum. Eash, Rasher and Waxman (1979) have included 'Variability in teaching methods and materials' as an important dimension in an instrument developed by them for evaluating teacher behaviour. Taylor and Colin (1979, p.20) and Borich and Madden (1977, p.134) have also recognised the important role played by 'Teaching methods and procedures' in influencing the effectiveness of a curriculum.

These observations provided the basis for selecting 'Instructional methods and procedures' as a hypothesized factor.

5. Task Orientation of the Class

Walberg and Moos (1980, p.67) have identified Task orientation as an important dimension of instructional environment. According to them 'task orientation' refers to emphasis on achieving specific academic activities. Trickett and Moos (1973) have included 'Task orientation' as one of the dimensions in their classroom Environment Scale. Similarly, Fraser, Treagust and Dennis (1984) have also included this dimension in the College and University classroom Environment Inventory developed by them. On the basis of this rationale
'Task orientation of the class' has been selected as a hypothesized factor.

6. **Order and Organisation of the Class**

   According to Welberg and Moss (1980, p.67) Order and organisation of the class— which refers to the extent to which the instructional environment is orderly and clear in its expectations, maintains control and responds to change— is an important factor influencing the effectiveness of the curriculum. In the Learning Environment Inventory (Anderson and Walberg 1974a) 'Disorganization' has been included as a dimension. Trickett and Moos (1973) have included 'order and organization' as one of the dimensions in their Classroom Environment Scale. Washton (1967) has pointed out the use of items measuring order and organization under the variable 'Classroom management'. The work of the researchers cited above justified the selection of 'Order and organisation of the class' as a hypothesized factor.

7. **Evaluation Procedures Used**

   Davis, Alexander and Yelon (1974, p.113) have stated that 'Grading and Examinations' followed in a course influence the instructional effectiveness of the course. Taylor and Colin (1979, p.116-117) and Wittrock (1970, p.7-9) have also recognised that procedures used for evaluating students performance influence the effectiveness of the operational curriculum. Becker, Geer and Hughes (1968) have established the powerful influence of assessment procedures on student learning in a study carried out by them at Kansas University. Based on these studies 'Evaluation Procedures used' has been included as a hypothesized factor.
8. Feedback Provided to Students

Davis, Alexander and Yelon (1974, p.112), Oliver (1977, p.310) and Simpson (1969, p.87) have all highlighted that provision of appropriate and adequate feedback to students influences the effectiveness of instruction and hence the achievement of curricular objectives. Walberg and Thomas (1972) have emphasized the importance of diagnosing the learning difficulties of students and providing guidance to them. These observations formed the basis for selecting 'Feedback Provided to students' as a factor hypothesized to influence the effectiveness of the curriculum under consideration.

Thus, based on the review of literature a total of nineteen factors have been hypothesized to influence the effectiveness of the chemistry curriculum under consideration.

2.9.0 SUMMARY AND DIRECTIONS

It is evident from the literature review completed that in India research on curriculum evaluation is still in its infancy. The review provided a deeper insight which enabled the formulation of an appropriate design to carry out the study. The following are the salient points brought out by the literature review.

1. Curriculum evaluation is concerned with those activities which provide information useful in course improvement and which show how effective the course is under specified conditions in the educational institutions.

2. Gain ratio which represents Gain as a proportion of possible
gain can be used as an index of the effectiveness of the curriculum.

3. The best measure of effectiveness of a curriculum is the percentage of students achieving its objectives.

4. When the curriculum to be evaluated is vast as is the case in most of the courses, domain sampling technique may be adopted to arrive at a representative sample of the curriculum in terms of objectives.

5. Agglomerating many types of post-course performance into a single score is a mistake, because failure to achieve one objective is masked by success in another direction.

6. For several obvious reasons norm referenced tests are considered unsuitable for the measurement of curriculum effectiveness. Criterion referenced tests are very much suited for this purpose. A criterion referenced test is used to ascertain an individual's status (mastery/non-mastery) with respect to a well defined behaviour domain.

7. The 'Operational Curriculum' deals with the processes of teaching and learning, Organization of the class and the milieu in which instruction takes place.

8. Classroom environment is a major determinant of learning outcomes. Therefore it is desirable to include classroom environment dimensions as independent variables in studies which attempt to evaluate effectiveness of curricula in terms of their impact on student learning.

9. Classroom observation schemes focus upon 'low inference' measures of classroom environment; but student and teacher perceptual measures focus on 'high inference' variables.
10. Student perceptions of their classroom environments account for appreciable amounts of variance in learning outcomes often beyond that attributable to background student characteristics such as pretest performance, and general ability.

11. Although a large number of instruments are available for assessing the psychosocial aspects of classroom environment, very few instruments are available for assessing the instructional aspects of classroom environment.

12. The studies reviewed enabled identification of a number of factors related to (i) the student, (ii) the teacher and (iii) the instructional environment, which are expected to influence the effectiveness of the chemistry curriculum of polytechnics.

The design of the study is described in the next chapter.