Unesco-UNEP International Environmental
Education Programme

Environmental
Educational Series 12

Evaluating Environmental Education in Schools

A practical guide for teachers

prepared by
Dean B. Bennett

Division of Science, Technical
and Environmental Education

The opinions expressed in this publication are those of the authors and do not necessarily coincide with any official views of Unesco. The designations used and the presentation of the material herein do not imply the expression of any opinion whatsoever on the part of Unesco concerning the legal status of any country, or of its authorities or concerning the delimitations of the frontiers or any country or territory.

© Unesco 1984, reprint 1989
ED 84/WS/77

FOREWORD

The various international and regional meetings on environmental education (EE) organized since 1975 by Unesco, with the co-operation of the United Nations Environment Programme (UNEP), and particularly the Intergovernmental Conference on Environmental Education (Tbilisi, USSR, 1977), have stressed the importance of elaborating guidelines and instruments which would stimulate those in charge of educational activities at various levels to develop an environmentally related pedagogy.

In this perspective, the Unesco-UNEP International Environmental Education Programme (IEEP) has placed special emphasis on the development of contents and methods which would favour efficient integration of EE into school and out-of-school educational processes.

Among the principal methodological questions related to EE, the evaluation of educational practice appears to be essential. In fact, evaluation should provide useful information about changes that should be made in the conceptions underlying the preparation of educational and training programmes, and about how teaching materials can be improved. However, the evaluation of environmental education requires more than the approaches, methods and techniques that are commonly used in the evaluation of general education. In addition to cognitive achievements, evaluation in environmental education must appraise competence in decision-making, problem-solving and the organization of action, as well as the values which determine the orientation of individuals and groups towards the environment.

The complex aspects involved in EE evaluation explain its limited development to date in most countries. Some recent attempts to produce better EE evaluation tools have been made in connection with certain actions undertaken by IEEP, namely the instruments proposed in modules related to the management of natural resources (Environmental Education Series No. 3) and to environmental problems in cities (Environmental Education Series No. 4). They remain however limited to specific areas or problems.

This guide to evaluation of EE in schools has been prepared by Dr Dean Bennett, professor at the University of Maine at Farmington and well-known international specialist in environmental education, and in the framework of IEEP. It attempts to meet pressing needs in this field by proposing a general approach to EE evaluation as well as practical knowledge which can be handled by the teachers themselves.

The guide comprises six chapters which address, step by step, different questions to be considered in a truly scientific evaluation process, though in a practical and accessible manner. Evaluation and associated statistical instruments which involve a certain complexity are given in the appendix.

In its original development and organization, the guide is experimental in character. Readers from various countries should consider both content and approaches in a critical perspective, examining suitability to local conditions and/or proceeding to the necessary adaptations before using them.

*

* *

The opinions expressed in this publication do not necessarily reflect the official position of Unesco. The designations employed and the presentation of materials throughout this publication do not imply the expression of any opinion whatsoever on the part of Unesco concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

TABLE OF CONTENTS

PREFACE

Teaching is a lot like running a marathon - once committed, it takes great amounts of energy just to finish. But when it's over and a little rest and time have diminished memories of aches and pains, thoughts of the next race are entertained and an analysis of what happened invariably takes place: the opinions of observers are sought, personal feelings recounted, and the experience reconstructed. Gradually a plan takes shape and anticipation of the next event develops.

This is a guide to help you, the teacher, not only run the last mile but do it better each time. It will help you reach your environmental education goals more efficiently and effectively. The procedure presented here is clear and simple - four easy steps:

1. deciding what to evaluate;
2. planning how to do it;
3. carrying it out;
4. using the results.

The directions for each step and the alternative pathways and routes to accomplish it are based on the assumption that the teacher is the evaluator. Evaluation is not something everyone likes to do, too often it is avoided, except to assign grades. But there is in most teachers, a stir of curiosity - 'Is what I'm doing really worth while? Is it the best I can do for my students' - because, when it comes right down to it, teachers care about what happens in their classrooms.

The methods of evaluation presented here will help you find out what is happening in your environmental education programme so that you can improve your teaching, help your students grow, and make our environment a better place to live. You will find methods to help you plan your programme and anticipate its results. Other methods are based on the view that your programme is a happening, and because you can't predict everything that will occur, you must probe to discover the unexpected, the unanticipated. There are methods to help you measure what happens objectively, and there are others to help you recognize results less easily measured - requiring more reliance of your intuitive sense.

Following the chapters detailing the steps in programme evaluation, Chapter VI presents a case-study to help you see more clearly how evaluation can be used to strengthen the role environmental education in a school's curriculum. Chapter VII, the final chapter, provides sources of additional information and help.

The preparation of a book, such as this, cannot be done alone, for we stand on the shoulders of others. A very special thank you is given to Ian Robottom, Chairman, Research Group on Curriculum Innovation and Evaluation, Deakin University, Victoria, Australia, not only for his review and valuable critique of the outline and entire manuscript, but for his views of evaluation in environmental education - views which influenced greatly the usefulness and practicality of this guide. The reviews, thoughtful critiques, and words of encouragement from Harold R. Hungerford, Professor, Department of Curriculum, Instruction and Media, Southern Illinois University at Carbondale, and George T. O'Hearn, Professor and Director, Office of Educational Research and Development, the University of Wisconsin at Green Bay, are also acknowledged with deep appreciation.

Gratefully acknowledged are the many helpful suggestions of L. Albala-Bertrand, Programme Specialist, Environmental Education Section, Unesco. A special thank you is given to Sylvia Hodgkins for her extraordinary effort, unusual care and attention to detail in the typing of the manuscript. Finally, to my wife, Sheila, I owe yet another debt of gratitude for her constant encouragement, inspiration, insight and patience for the hours we lost together.

Dean B. Bennett
Mt. Vernon, Maine
United States of America

May 1984

INTRODUCTION: WHAT IS EVALUATION AND WHY SHOULD I BOTHER WITH IT?

WHAT IS EVALUATION?

Should you be concerned with evaluation? Do you want your students to be more caring of their environment - capable stewards? Do you want to use your limited time, money, and other resources wisely? AND DO YOU WANT TO BE ABLE TO MAKE YOUR OWN JUDGEMENTS AND DECISIONS IN THESE MATTERS? If you answered 'yes' to these questions, this book is for you.

Two keys to outstanding teaching are effectiveness and efficiency. And this book is your guide to both. Effective means that the effects or results of your efforts are worth while, for example, your students achieve a high level of proficiency in critical thinking concerning environmental issues. Efficient, on the other hand, means that the methods you use and the settings you create are the best in terms of their cost, time required, and appropriateness to the task, for example, the time and money you have to teach environmental education is limited, so you use to advantage every minute of a field trip.

Evaluation, then, as used here, is judging the worth or value of your environmental education programme - its products, ends, or outcomes (its effectiveness) and its processes, means, or ways of doing its job with limited resources (its efficiency). Evaluation differs from research, which is not so much concerned with making judgements as it is with searching, seeking out, and gathering new information to further knowledge.

WHY IS EVALUATION IMPORTANT?

The importance of evaluating your environmental education teaching is found in the direct, tangible benefits that will come to your students, your school, and, above all, to you. Not only will you become a more efficient teacher, but you will be seen as an educator who wants to improve and does something about it. By the mere act of evaluating you will demonstrate accountability.

Perhaps one of the best ways to view the benefits from evaluation is to see them as four interrelated components of your environmental education programme (see Figure 1).

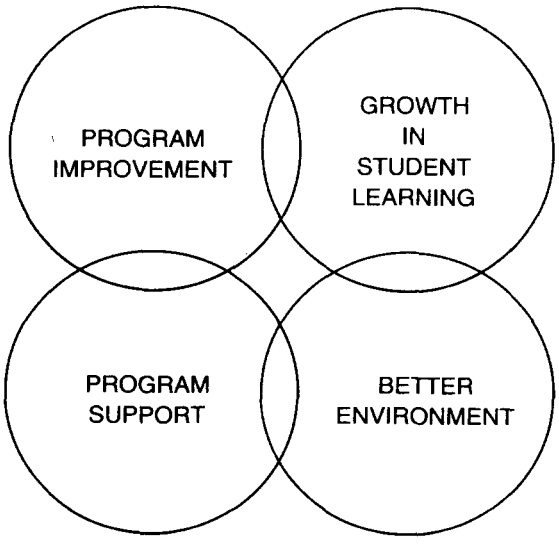


Figure 1

COMPONENTS OF AN EVALUATION PROGRAMME

Evaluation Can Lead to the Improvement of your Instructional Programme

First, evaluation can result in the improvement of the effectiveness and efficiency of your teaching methods and learning activities.

Second, evaluation can result in the improvement of the effectiveness and efficiency of the learning environment. This includes your classroom, the environment outside of the classroom - the physical, emotional, and social climate of the school setting and the community, and the instructional resources at your disposal.

Evaluation Can Lead to Greater Growth in Learning by your Students

This, of course, is a goal of paramount importance. Growth in learning as used here means the acquisition of knowledge, the clarification of values and development of moral reasoning, and the improvement of independent critical thinking and action skills. Evaluation of these important dimensions can help you in three ways: (1) diagnosing the learning needs of your students so that you can be more efficient and effective in correcting deficiencies and fostering growth; (2) measuring achievement so that you can assess the efficiency and effectiveness of your programme; (3) using evaluation, itself, as an effective teaching tool.

Evaluation Can Lead to a Better Environment

The ultimate goal of environmental education is a healthful and healing environment. Learning by doing is a major way of teaching environmental education. Students can become involved in projects in which they act directly to improve their environment, such as, tree planting. Students can also become involved in projects in which they act indirectly to improve their environment, such as, communicating a concern about an environmental problem to an official who can then act directly on it. Evaluating the environmental effects of these activities can help you judge not only the growth in your students but the worth of your instructional programme.

Evaluation Can Lead to Greater Support for your Programme

As stated earlier, the mere act of evaluating indicates to others that you care about your students, the job you are doing, and the environment of which you are a part. And as you acquire and communicate evidence of the success of your efforts, you will gain the respect and support of school personnel and the general public. But it's your students who will be the direct beneficiaries of your programme and who will see first-hand its results. They will be, perhaps, your best supporters and ambassadors.

HOW DOES EVALUATION WORK?

One way to think of evalution is to picture a wheel with evaluation as the hub (see Figure 2). The rim of the wheel is the educational process involving the continuous cycling of interrelated events: assessment of student needs, consideration of goals and objectives, development of the instructional activities and learning environment, and growth in student learning. The spokes are, in effect, the links between the hub and rim and support the rim at its four key points. If each point is not kept in good adjustment and balance with the others, movement of the educational process could be somewhat wobbly and out of line.

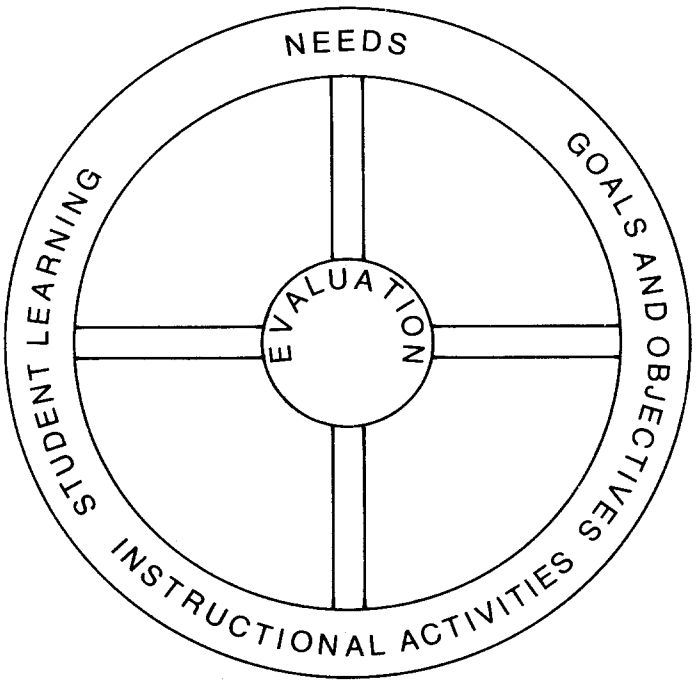


Figure 2
EVALUATION AND THE EDUCATIONAL PROCESS

HOW CAN THIS GUIDE HELP ME?

Building a wheel and keeping it in good repair is the job of a wheelwright, and a good wheelwright knows the importance of the hub and spokes. This is where you come in, because in this guide it is you, the teacher, who is the wheelwright. You have the responsibility for building and keeping your instructional programme in good repair and rolling smoothly in the right direction. As you form your evaluation programme, you will be using formative evaluation whenever you check on your progress. Determining the performance of the finished product is called summative evaluation - looking at the sum total of your efforts.

YOU AND YOUR STUDENTS AS EVALUATORS

As already discussed, this guide is based on the view that you, the teacher, and your students should be intimately involved in the evaluation of the environmental eduation programme. After all, you and your students are the ones most closely associated with the programme, who spend the most time with it and have the most influence on it, as well as being the ones most influenced by it.

It is you, the teacher, who is most directly responsible for your programme's efficient and effective operation. Your role in evaluation, therefore, is simply to gather information and make decisions based on that information - decisions that will ensure the efficiency and effectiveness of your programme.

But not only is it important for you, the teacher, to be an evaluator, but so should your students be evaluators. As recipients of the programme, they have a rightful role, and their shared opinions, feelings and insights about the programme, what they learned and the evaluation, itself, can be of special value.

A SPECIAL CONSIDERATION REGARDING THE ROLE OF STUDENTS[1]

There is yet another, perhaps even more important, reason why students should be involved in evaluating programmes of environmental education. As will be discussed in the next chapter, central among the purposes of environmental education, is the development of independent critical thinking. If it is agreed that this is, indeed, an important goal, then you must be alert to the possibility of your evaluation processes working in contradiction to this goal. For example, if you have not established an atmosphere free of censorship and one in which they perceive that their judgements will not work against them, your students will not be encouraged to think freely and critically about environmental problems and issues. Evaluation, by its nature, suggests to your students that judgements will be made about their performance. You can be sure that in most cases, they will be aware of this. It is then incumbent on you to incorporate measures in your evaluation programme that give consideration to the following suggestions:

1. make purposes of your evaluation programme and use of the results clear to your students and non-threatening;

2. give students every opportunity to participate in the planning of the evaluation programme;

3. consider the rights of your students with regard to the access and release of information;

4. ensure anonymity in release of the information.

OVERCOMING OBSTACLES TO EVALUATION

Mystery

By now you should know that there is no mystery to evaluation. It is simply gathering and using information to judge the value of whatever you are doing. In other words, it helps you decide how efficient or effective your efforts are and how they can be maintained and improved.

Time

Perhaps the greatest obstacle to evaluation is making time for it. The secret to overcoming this constraint is to build evaluation into your programme from the

beginning, budgeting time for it just as you do for preparation and teaching. In fact the testing of students is an area of evaluation which can in itself be a teaching tool.

Sometimes, it is also possible to spread out your evaluation activities by concentrating at different times on different components of your programme such as classroom setting and various teaching methods. In other words, if time is limited, don't do it all at once; take a long-range view.

Another recommendation is to limit your evaluation to what is most important, for example, improving your programme, increasing the understanding of a difficult concept or skill, or gaining support. Focus your efforts. Determine your priorities.

Expense

The major expense in most evaluation is in the hiring of a consultant. This guide is designed to help you become the evaluator. It is based on the premise that a worthwhile evaluation can be conducted by the teacher. Of course, there are times and places in the process where outside help would be of benefit and these are pointed out. But, practically speaking, you are capable of carrying out the important steps by yourself.

Subjectivity

One criticism you will undoubtedly encounter relates to subjectivity - how can you, the teacher, be objective about your own programme? Will you be honest in your appraisal? Will your results be believable? Won't your biases show in the questions you develop, your analyses and conclusions?

The fact is, however, that one important goal of evaluation is to show you how to improve the efficiency of your own efforts - to make your job easier. Therefore, because you plan to use the results practically and will benefit personally, it is in your interest to be sincere and honest. And since evaluation is also designed to help your students grow and develop and ultimately to help create a better environment, its effects are open for all to see and are subject to outside scrutiny and discussion. Finally, if you openly invite inspection and review, seek reactions and opinions from students, parents and the community, and report your findings honestly, you can overcome much criticism of bias in your approach and interpretation of findings.

Complexity

If you clearly define what you want to accomplish through evaluation and follow the four easy steps in evaluation listed below, you will successfully remove much of its complexity. This guide is organized to help you carry out the four steps easily with a clear understanding of the decisions you will need to make.

FOUR EASY STEPS

The following steps form the framework for your evaluation:

1. deciding what's important to evaluate;

2. planning the evaluation;

3. conducting the evaluation;

4. making use of the results.

USING THIS GUIDE

The remainder of this guide is directed towards helping you to clearly understand and carry out each of the four steps. Together they can help you improve your programme and its results. This guide is designed as a reference and a set of directions to help you either develop a comprehensive evaluation of your total programme or develop an evaluation of some specific aspect of your programme, for example, your goals and objectives, special activities, or instructional resources.

To become acquainted with its possibilities, it is suggested that you first skim each chapter noting headings and illustrations. For an overview of how evaluation of a programme works, you might first read Chapter VI: A Case-Study of an Evaluation. Chapter VII provides sources of additional help useful in solving specific problems.

CHAPTER I: WHAT SHOULD I EVALUATE? (STEP ONE)

INTRODUCTION

The question, 'What should I evaluate?', is central to all evaluation. If you are a typical teacher of environmental education, you have limited time, money, and expertise in evaluation. Your situation, then, may not be too far from that of an aged cat stalking an unseen prey for one more meal - you've got to make everything count! And, in particular, it's important to know what you're after:

Are you most concerned about what has happened to your students because of your teaching? For example, is there evidence that they do not have a grasp of basic ecological concepts or cannot think critically about environmental issues?

Are you most concerned about the effectiveness and efficiency of various ways you teach? For example, do you suspect that more involvement in real environmental problems will produce better results?

Are you most concerned about the influence of different aspects of the learning environment on the results you obtain? For example, do you feel that you should make greater use of an outdoor classroom?

Are you most concerned about the effects of your programme on other people and the environment? For example, are there environmental problems or issues, such as the deterioration of a local park or lack of a playground, your students would like to do something about?

Are you concerned about the appropriateness of your goals and objectives to meet real needs?

Are you interested in how the results of your programme compare with other programmes?

Deciding among these questions can be confusing, but looking at the needs and goals of environmental education can help set priorities.

NEEDS AND GOALS OF ENVIRONMENTAL EDUCATION

Fundamentally the need for environmental education arises from the need for an environment fit for life - a healthful and healing environment. The need to protect and maintain such an environment is a human responsibility, and the role of education, under the name environmental education, has received much attention in recent years as one means of meeting this responsibility.

Many goal statements have been written that reflect this responsibility. One, on which there is widespread agreement, is from the Belgrade Charter: A Global Framework for Environmental Education.

'The goal of environmental education is: to develop a world population that is aware of and concerned about, the environment and its associated problems, and which has the knowledge, skills, attitudes, motivations and commitment to

work individually and collectively towards solutions of current problems and prevention of new ones'.[2]

One important value of this kind of goal statement lies in the direction it can provide to teachers in setting goals for their own programme. For example, the goal of the Environmental Education Project for the schools of Guam is described as follows:

'To graduate citizens who are knowledgeable and conscientious about the environmental problems of Guam and the world'.[3]

This is a practical example of a goal established for a specific programme.

Every educational programme is based on perceived needs and has goals responsive to these needs. Needs provide the rationale for a programme, and its worth is assessed in terms of how well its goals, objectives, activities and results meet the needs. As an evaluator of your programme, you should frequently assess the relevance of your programme's goals to what you consider to be the important needs. Do your goals, for example, address all the major outcomes you hope will result from your programme? Keep yourself open to the desirability of new goals as you uncover new needs through your programme work and evaluation activities.

DECIDING WHAT ASPECTS OF STUDENT LEARNING TO EVALUATE

The role of the human in maintaining a quality environment is central to most goal statements in environmental education. It is, therefore, no surprise that what happens to the learner as a result of a programme is also a central question of most evaluation efforts. Finding the answer helps predict how students might act or not act with respect to their environment, and it can often suggest what aspects of the environmental education programme are most or least effective.

Objectives

One school of thought believes that evaluation must be based on determining first what you want to happen at the end of the programme. In other words, one should establish a clear set of objectives. Another school believes that education is fundamentally a happening - an unpredictable and unreplicable experience, full of surprises and spontaneous events, and evaluation should be a let's-see-what-happened investigation. In this latter case, the programme is essentially viewed as goal free.

This guide is based on the belief that both approaches have their merits, therefore, a variety of methods of evaluation are presented - some that use predetermined objectives and others that rely less on specifying outcomes in advance. How you balance these two approaches is one of your first decisions. The recommendation here is that you be frugal in the number of objectives you establish, allowing opportunities for unanticipated outcomes to occur and time to search for them.

Objectives differ from goals in that they are more specific, not as long-ranged, and more measurable. Frequently programmes use two levels of objectives: (1) those that are general, non-performanced-based and more subjectively measured; and (2) those that are specific, performanced-based, and more objectively measured.

General Objectives

General objectives are useful in describing outcomes in broad terms and providing direction for more specific and measurable behavioural objectives. Some programmes, for example, have derived three general objectives from the goals presented earlier: a cognitive objective, an affective objective, and a skill objective. If you wish to use general objectives, the following examples may give some guidance:

* General cognitive objective:

To help students acquire an understanding of their global environment, their relationship with their environment, and environmental issues and problems.

* General affective objective:

To help students acquire an appreciation of their environment and an active concern and motivation to maintain and improve the quality of their environment.

* General skill objective:

To help students develop thinking and action skills for the prevention and resolution of environmental problems and issues.

You will notice that the above objectives are general enough to give both direction and flexibility to your teaching. From the standpoint of evaluation, you are free to search out and accept a wide range of evidence that the objectives have been met. The use of such terms as appreciation, interest, motivation, understanding and thinking and action skills, without specifically defining them in observable actions or behaviour, encourages the use of subjectivity and intuition in assessing your programme's outcomes. Some believe that this approach increases the possibility that you will look at the programme from a broader perspective, conduct a greater search for unanticipated outcomes, and be more apt to consider outcomes not easily measured - such as attitudes and values. Others find it difficult to accept this level of assessment; they argue that you should only accept what you can observe and measure as direct evidence for the accomplishment of objectives. They are the ones who promote the use of behavioural objectives, to be discussed shortly.

A World about Objectives and Quantitative and Qualitative Approaches

The following section on behavioural objectives emphasizes quantitative approaches to evaluation. As you will see, by its nature, a behavioural objective seeks to measure and quantify behaviour. The amount of space devoted to describing these objectives and their development does not necessarily imply that they are more important than the qualitative approaches emphasized in the sections on intuitively measured outcomes and unanticipated outcomes. Qualitative approaches, for example, diaries, questionnaires, and interviews, can yield data describing the nature, characteristics, properties, attributes of a student's attitudes and abilities. One final note: it's a rare objective that is not associated in some way with both quantitative and qualitative data, and most of the behavioural objectives that follow have the potential of yielding qualitative information.

Behavioural Objectives

Behavioural objectives are simply statements of how you would like your students to behave or perform as a result of your teaching - what you would like them to do that indicates they have grasped a concept, acquired an attitude, or developed a skill. 'Do' is the key word, meaning something the student does that is observable and measurable.

In a behavioural objective (also called a performance objective), what the student will do is described by an action verb, carefully selected to indicate the likelihood that the student possesses the desired values and attitudes, knowledge and understanding, thinking skills, and action skills. Here is an example: 'Given a set of pictures of environmental problems and another set showing people engaged in activities associated with causes or problems, the student will correctly match 75 per cent of the problems with their probable causes'. Notice that the word 'match' is the action verb. The example also includes the two other common components of behavioural objectives: (1) the conditions under which the student will demonstrate the desired outcome, i.e. given a set of pictures; and (2) the extent or degree of mastery, i.e. 75 per cent accuracy. Incidentally, in this example, by selecting different sets of pictures, you could vary your test to assess different outcomes, for example, knowledge and understanding of ecological concepts, the higher thinking skills of analysis and evaluation, or even, to an extent, values and attitudes.

When you develop or select objectives, you can do so at two different levels - one at the student level and the other at the class or group level.

* Objectives at the student mastery level

Objectives in this case give percentages that relate to how accurate the individual student should be in responding to a test. The objective given above is a good example: 'Given a set of pictures of environmental problems and another set showing people engaged in activities associated with causes of problems, the student will correctly match 75 per cent of the problems with their probable causes'. Here your concern is with the growth of each student, and, the objective gives you a basis for diagnosing individual needs.

* Objectives at the class mastery level

Objectives, in this case, give percentages related to the achievement of the class as a whole and give results appropriate to evaluating the total programme. An example, using the previous objective, would be:

> 'Given a set of pictures of environmental problems and another set showing people engaged in activities associated with causes of problems, 70 per cent of the students will correctly match 75 per cent of the problems with their probable causes'.

One important point related to this kind of objective concerns the setting of standards, say 70 per cent or 75 per cent. These are a matter of judgement and should reflect a local level of achievement that is publicly acceptable, reasonable and justifiable. Because these objectives establish criteria for determining levels of student growth, tests based upon them are frequently called 'criterion-referenced tests'.

In deciding between the use of general objectives or the more specific behavioural objectives, it is recommended that consideration be given to using both. You will undoubtedly have a few specific, measurable outcomes in mind which may

be appropriately stated as behavioural objectives. In this case, be judicious in your selection and develop only a very few that have the highest priority. To keep yourself open to unanticipated outcomes and those less measurable, also consider using a few general objectives.

Finally, you should always be concerned about the worth of your objectives, themselves. Constantly question their value and be quick to modify or change objectives if you have evidence that they are inappropriate. As you collect and analyse data from your observations and tests, you will often see the practicality of your objectives in new light.

With this general background about objectives, you are now ready to consider the evaluation of four key dimensions of your students relating to affective, cognitive, and behavioural areas of development.

Values and Attitudes

Without question the values we hold play a central role in problems of the environment. In fact, it may be said that issues are conflicts in values. The Unesco-sponsored Intergovernmental Conference on Environmental Education (Tbilisi, USSR, 1977) recognized the importance of values in its report: ' ...All decisions regarding the development of society and the improvement of the lot of individuals are based on considerations, usually explicit, concerning what is useful and beautiful and so on'.[4]

Values, then, as defined here, relate to the ways we judge the worth of things and are thought to be so powerful that they have a pervading influence over our lives and guide our very actions. Perhaps this is because we do our best to protect and keep things that we value.

Values, for our purpose, derive from attitudes, which, like values, also reflect our feelings towards things but include a wide range of emotions that influence the extent to which we value something. For example, attitudes of dislike and fear of toxic chemical waste can strengthen the extent we value a clean environment.

Attitudes and values are always directed towards tangible objects and events or intangible ideas and thoughts. As an environmental educator, you should consider the development or clarification of attitudes and values towards the following:

nature

human relationships with nature

built environment (human design and development)

oneself (individual self-image)

others (including future generations)

Developing Measurable Objectives

Measuring attitudes and values is the most difficult task in evaluation because they cannot be measured directly. For example, a student may tell you that he or she values the conservation of our scarce energy resources but in reality does not want to give up an energy-wasting lifestyle. How can you know for sure? The fact is, you probably won't, but there are ways that will help you obtain a better idea. Most evaluators agree that the greater the number of different approaches used to assess an attitude or value, the greater the likelihood that you will increase your certainty of the extent an attitude or value is held.

What are some approaches you can use? Direct observation of student activities, for example, looking for the number of students volunteering to clean up a park or participate in a recycling project, is one approach. Or artefacts, like student diaries, might yield useful information, for example, how effective students feel they can be in preventing environmental problems. You can be even more certain of student attitudes and values if you observe students when they are not aware of being evaluated. For example, you could find out the kinds of books being checked out of the library. Some have even gone so far as to use a contrived situation to test student reactions, for example, purposely placing litter along the route of a field trip on the school site to see how much is picked up. Other approaches use pencil and paper tests.

The important thing to remember when you decide how students will demonstrate their attitudes and values is that they do so in situations free from outside influences so that you can see their true feelings expressed. Note, also, that in all the suggestions given, qualitative information, as well as quantitative data, may be obtained:

Some action verbs you can consider using are:

volunteer

freely express

freely choose

willingly do

spontaneously respond.

Below are some sample behavioural objectives reflecting these approaches. The examples chosen are concerned with the value of wildlife habitat. Notice that objectives one and three, in particular, provide opportunities to gather qualitative information about the nature of students' feelings.

1. After hearing an announcement expressing concern over the loss of nesting sites of a particular bird, 50 per cent of the students will volunteer to join a group working to correct the problem.

2. As a result of a unit of instruction on wildlife, the number of books related to wildlife taken out of the library will increase by 25 per cent.

3. Following a two-week programme on the programme of diminishing wildlife habitat, one-third of the students will select to prepare a research paper on wildlife habitat improvement from a list of five possible areas of study.

4. At the end of a week-long study of an endangered species and what people can do to help, 40 per cent of the class will indicate on a scale of one to five that they are confident or strongly confident of their ability to actually help contribute to a solution of the problem.

5. Following a course on wildlife management, 60 per cent of the students will indicate on a five-point scale a strong or very strong concern for the loss of habitat.

6. Following a study of specific kinds of natural areas, 75 per cent of the students will indicate that they consider wetlands valuable habitats for wildlife by placing a mark on a line between the two terms worthless and valuable.

Knowledge and Understanding

Knowledge of something means knowing about it and being able to recall it from memory. Understanding something is at a higher cognitive level and means comprehending its meaning and being able to explain it.

It is generally agreed that preventing and solving environmental problems requires a broad knowledge and understanding of many things including concepts related to ecology, engineering and design, economics, government, sociology, psychology and human health. The question you will need to answer is what do I want my students to know and understand among the many possibilities? In other words, what important facts, concepts (ideas), principles, and generalizations should I teach (and evaluate)? Your goals and general objectives will get you started in the right direction. The following categories might also help you to avoid leaving out important facts or concepts:

nature and ecological systems

natural resources

the built environment

environmental problems and issues

social systems - political, economic, legal

arts and humanities

social interaction

individual or self-development

dimensions of space and time.

Whatever you target as key for students to know and understand, you should be convinced of its importance. Choose wisely.

Developing Measurable Objectives

The evaluation of what your students know and understand is much easier and more direct than assessing attitudes and values. A knowledge of the effects of air pollution can be tested, for example, by having students simply recall what they know; and understanding, which is a higher cognitive level, can be tested, for example, by having the students explain the relationship between air pollution and human health. Because the assessment of knowledge and understanding is more direct, fewer measures of the same understanding are needed.

When writing behavioural objectives, the following action verbs are suggested to describe how students could demonstrate knowledge and understanding in environmental education:

recall

repeat

state

define

tell

explain

list

duplicate

give examples.

Following are examples of behavioural objectives:

1. When given the components of an ecosystem, such as a wetland, 75 per cent of the students will be able to correctly connect them into a food web.

2. When presented with a problem of resource scarcity, the student will be able to explain its relationship to the concept of exponential growth.

3. Given a series of ten slides of environmental conditions, 70 per cent of the students will correctly check the most probable cause of eight of them.

4. Given the task of researching an environmental issue, 60 per cent of the students will compile a list containing at least ten different sources of information.

Thinking Skills

According to the Tbilisi conference on environmental education, the problem-solving approach is probably the most important distinguishing characteristic of environmental education. The report of the conference went on to detail the questions individuals need to ask if the approach is to be used effectively? Who makes decisions affecting the quality of our environment? Why were they made? According to what criteria? Were long-term consequences considered?

These and other similar questions are at the root of the critical thinking required to prevent and resolve environmental problems and issues. But these higher level thinking skills are more difficult to develop and assess than knowledge and understanding. And you should be aware that they will, more than likely, require greater amounts of time. However, in view of the special contributions of environmental education to the world society, the time and energy you devote to developing these skills will be well spent.

Today there appears to be wide agreement that there are different levels of cognitive skill progressing from knowledge - the lowest level - through comprehension, application, analysis, synthesis and evaluation - the highest level.[5] Each level of thinking includes the skills below it. As a teacher intent on achieving the goals of environmental education, you will need to help your students acquire the higher thinking skills that lie at the heart of critical thinking and problem-solving.

Developing Measurable Objectives

To help you plan the development and evaluation of critical thinking skills, each is treated briefly below.

1. Application is the ability to apply knowledge and new ideas to developing better understanding and solving problems.

   Action verbs that describe what your students can do to demonstrate they have acquired the skill of application are:

apply	devise a method
prove	plot a graph
organize	reformulate

A sample behavioural objective:

Given a set of guidelines on the art of testifying at public hearings, 75 per cent of the students will receive a rating of eight on a ten-point scale assessing effectiveness.

2. Analysis is the ability to break down ideas, objects, and events, including problems, and issues, into their parts. In so doing underlying assumptions, causes, effects, reasons, etc., are revealed and their organization becomes clear.

Action verbs that you can use in developing analysis behavioural objectives include:

detect	identify
compare	discriminate
contrast	distinguish
analyse	specify

A sample behavioural objective:

While reviewing statements in support of and opposed to an environmental development, 75 per cent of the students will be able to detect bias and faulty assumptions.

3. Synthesis is the ability to put parts and elements into a unified whole. For example, after analysing the causes of an environmental problem, a workable solution might be created that takes into account the causes of the problem.

Action verbs that reflect the skill of synthesis include:

create	reorganize
predict	generalize
integrate	plan

A sample behavioural objective:

Given an understanding of the elements of good design, 75 per cent of the students will create a plan, meeting minimum criteria, for beautifying the school grounds.

4. Evaluation is the ability to judge the value of ideas, objects, and events using appropriate criteria. For example, identifying an environmental problem, such as the severity of a water pollution problem, may require the evaluation of data from a water monitoring programme.

Action verbs that demonstrate the skill of evaluation include:

evaluate
judge
discuss critically
defend reasons
develop and apply criteria.

A sample behavioural objective:

Given criteria for judging the suitability of soil for a given land use, 75 per cent of the students will be able to evaluate its compatibility.

Action Skills

Not only is the problem-solving approach of environmental education perhaps its most distinguishing feature, but it's how this approach is used that often sets environmental education apart from other educational methods. Involvement with real problems and issues is seen as the most effective way to develop the action skills needed to investigate, evaluate and implement solutions to problems. Hungerford and Litherland identified six categories of environmental action: (1) persuasion; (2) consumerism; (3) political action; (4) legal action; (5) eco-management or direct environmental improvement activities, such as landscaping; and (6) interaction - a combination of two or more of the categories.[6]

Developing Measurable Objectives

It almost goes without saying that action skills are easier to evaluate. Unlike values, which can only be inferred from behaviour, the existence of action skills may be directly observed. When writing a behavioural objective for an action skill, you don't have to think about what kind of behaviour will demonstrate a particular skill because the behaviour you describe is the behaviour you hope will result. The only inference you need to make concerns the extent students have the thinking skills associated with the action skill. For example, an effective debate requires application of concepts, analysis, synthesis, etc., and trouble-shooting a piece of equipment requires analysis and a variety of other thinking skills. The extent these thinking skills are developed is a matter of indirect observation.

In the case of action skills, the action verbs you choose for your objectives are the skills you hope to develop. Here is a list of examples:

persuade	testify	lobby	map
boycott	write	monitor	photograph
debate	plant	calibrate	illustrate
dissect	operate	demonstrate	position

In evaluating skills, you will also need to consider efficiency (performing with the least amount of energy and time) and effectiveness (accomplishment of what was intended). In the behavioural statement these considerations represent the extent or level of performance.

Below are some examples of action skill behavioural objectives:

1. Given the necessary equipment, 60 per cent of the students will be able to test for dissolved oxygen.

2. Using conventional drafting equipment, 57 per cent of the students will be able to accurately draw a map of the school site to a scale suitable for a 11" x 14" sheet of drawing paper.

3. In a school presentation, the students will be able to persuasively argue the point of view of the class as determined by a survey conducted before and after the debate.

Intuitively Measured Outcomes

Another way to measure growth in student learning is to examine your own feelings and insights. Intuition is the sense of something not really evident. We often have a sense of whether something is right or not - often without concrete reasons.

Is intuition important? There are those who will argue vigorously against the use of intuition in assessing programme effectiveness. They say it's not objective; it's not based on 'hard' data - data that is observable and measurable. But is all that we learn measurable? Can we measure with certainty student feelings, for example, what students like about the programme and why, what motivated them or did not, what motivations and interests they, themselves, acquired and why? We all have feelings, and, although some will say they are not 'hard' data, it may be argued that they often provide evidence which cannot be ignored - evidence that can often be quantified. For example, if 90 per cent of the students said that they would have the confidence to participate in the resolution of an issue related to energy use, this might be judged meaningful.

There are those who maintain intuition does have its place, that it is something you can fruitfully examine. In fact, you can develop objectives involving intuitive measures, for example, students will describe in writing their feelings about the programme. And you not only can probe the feelings of your students about what they learned, the programme, and the evaluation process, itself, but you can probe the intuitive feelings of others - parents, other teachers, administrators, community acquaintenances. Interviews work especially well in collecting this kind of qualitative data, but you should also consider using questionnaires, rating scales, diary entries, checklists, and written perceptions.

Allowing for Unanticipated Outcomes

A major criticism of behavioural objectives relates to their tendency to focus the teacher so much on the achievement of predetermined outcomes that there is no consideration of what else the student learned. It is not unreasonable to expect that some of the most important results in student learning are unanticipated. Most will agree that it is impossible to anticipate all outcomes, much less have the time to write objectives that cover all possibilities and test for them.

Therefore, to allow for unanticipated outcomes related to student growth, it is recommended that you carefully choose to write objectives for only a very few outcomes of the highest priority. This will give you room to search for other, sometimes more elusive and subtle, but no less important, quantitative and qualitative changes in affective, cognitive and action-skill areas. These might include new areas of career interest, improvement in reading ability, or the uncovering of an unknown talent. Sometimes learning problems are revealed such as a problem in working with another student. Once discovered, they can be dealt with. Looking for unanticipated outcomes requires the teacher to view children in their totality and be aware of all aspects of their development.

DECIDING WHAT ASPECTS OF THE INSTRUCTIONAL PROGRAMME TO EVALUATE

Evaluation is not only concerned with the improvement of the learner but with the improvement of the programme. Teaching is a process of constant refinement, of careful honing, designed to improve efficiency (accomplishing more in the same

amount of time) and effectiveness (improving its results - growth in student learning and, ultimately, a better environment). There are two major components of the instructional programme which are of special importance: (1) teaching methods and learning activities; and (2) the learning environment.

Teaching Methods and Learning Activities

Several teaching methods and learning activities characterize environmental education. The problem-solving approach is one of its most distinguishing characteristics.[7] Other methods and activities involve first-hand experiences with real problems and issues, including activities outside the classroom, values clarification and moral reasoning, interdisciplinary concept attainment, and critical thinking. Knowing how to select and apply these approaches efficiently and effectively is of crucial importance. It is to this end that evaluation of the instructional programme is directed.

Identifying Measurable Factors

Some factors of your selected methods and activities easily lend themselves to measurement: The amount of time spent in a particular activity, such as, student planning and building of an interpretive trail compared to growth in learning; the costs associated with particular activities; and the use and comparison of two different approaches, such as indoor vs. outdoor or lecture vs. inquiry to achieve mastery of the same concept or skill. Whatever you choose, the important thing is that the factors be observable, quantifiable and comparable.

Providing for Intuitively Measured Aspects

Here is where you would ask students what they liked best, what worked best, what they would like to do again - in short, how they felt about their experiences. These questions should be asked at various times throughout the programme so that changes can be made along the way. You should also ask yourself these same kinds of questions and be alert to your own feelings.

Allowing for Unanticipated Events

Too often a 'teachable moment' is missed because the teacher wouldn't or couldn't 'shift gears' to accommodate a sudden learning opportunity. It is, therefore, important to not view your methods and activities as the 'last word' in the art and science of teaching. One teacher, for example, started a week's unit on waste disposal and ended up with a six-week study of the problem of rats in the community - all because one student asked if rats at the town dump posed a problem. So be alert and flexible, and when you evaluate your programme, check to see if you are flexible in grasping teaching opportunities.

The Learning Environment

An important characteristic of environmental education is its focus on the real world, involving studies of natural and human ecosystems and associated problems. Environmental education, thus, takes place in many settings - the classroom, the school site, the community and beyond. How to best use these learning environments is the responsibility of the teacher. As a teacher you, of course, want to be able to control those important factors that influence learning in these settings. Knowing what they are and their influence will help you to effectively manage them.

Identifying Measurable Factors

* The classroom

In the classroom you have much more control of the setting. Here you have greater opportunity to create a learning environment - to improve its efficiency and effectiveness to accomplish your learning objectives. Important factors include:

comfort - temperature, humidity, light, sound

kind and arrangement of furniture

spatial relationships

design factors - colour, shape, form, texture, rhythm, repetition, emphasis

kinds and location of instructional resources - media materials, equipment, bulletin boards, books, etc.

The above are physical factors. More difficult to measure, but no less important, are psychological factors contributing to what has been called the classroom atmosphere or general mood of the learning environment. There are, of course, overt clues such as the kind and amount of questioning and discussion, comments exhibiting respect or lack of it, free expressions of happiness and humour, purposeful activity, time spent on task, and pleasant interpersonal communication. It is in assessment of the tone or mood of the setting that intuitive measures come into play. More about them later.

* The school and community outside the classroom

Outside the classroom you have much less control over the learning environment. However, you might be able to influence the design of the school site to enhance learning opportunities such as creating a pond for study, building a nature trail, or setting up soil management demonstrations. But generally, your major influence will be over the students' conduct - what they do, where they carry out their activities, and when. For example, you can control temperature by requiring clothing, conducting activities in shade, shelter, and other conditions at the study site, and planning activities according to the weather and time of day.

Generally, what has been said about the atmosphere in the classroom applies to the learning environment outside the classroom, except perhaps with one major difference: outside of the classroom you and your students will encounter a large amount of uncontrolled stimuli. You will need to focus attention. Studies show that time is needed for students to adjust, sort out, and direct their attention to subjects to be studied.

Another factor to consider relates to programme support. For example, is the school and community supportive of independent, critical thought and first-hand outdoor experiences? Without this kind of support, major learning opportunities will be limited.

Providing for Intuitively Measured Factors

Assessing the mood of a study, the atmosphere of the environment, the feeling of a place, essentially relies on your intuitive sense. It requires that you ask your students, as well as yourself, what is liked best, disliked and why; what should be retained, eliminated, and improved; what is comfortable and uncomfortable; what is aesthetic and unpleasing. Class discussions, interviews, and brief questionnaires can be used to gather perceptions, insights and feelings.

Allowing for the Positive Influence of Unanticipated Factors

A key to creating an effective learning environment is experimentation - trying new arrangements, introducing new materials and objects, being willing to make changes. Follow your hunches, listen to your students, probe to discover the effects of everything in the setting. If you do these things, you will be promoting a vibrant setting full of stimulation and richness - a place where learning is encouraged in unanticipated ways.

Deciding What Other Factors to Evaluate

An environmental education programme, by its nature, is focused on making education relevant to environmentally related concerns in the real world. Pursuing activities dealing with real problems and issues will influence both people and actual environmental conditions. Valuable information on the effectiveness of a programme can be gained by investigating the perceptions of people affected by the programmes and changes in the environment. How were views of people towards their environment and the programme changed? What do they suggest to strengthen the programme? What were actual environmental effects of student work? Will the results be long lasting? How do other people judge them? These are only some of the factors you should consider.

Now that you have a grasp of what the possibilities for evaluation are, you can turn your attention to planning a practical evaluation.

CHAPTER II: HOW CAN I PLAN MY EVALUATION? (STEP TWO)

INTRODUCTION

When you plan to evaluate your programme, above all, be practical. What can you, as a busy teacher, realistically do with limited time and resources? What is most important to you, your students, your administration, your community? What kind of information is most essential? Perhaps there are three key questions that form the basis for any programme evaluation plan:

(a) What are the questions of greatest concern as you look to improve your programme and accomplish your objectives?

(b) What are the best designs to use to answer the questions?

(c) What are the actual techniques and instruments you can use?

DETERMINING THE IMPORTANT QUESTIONS TO BE ANSWERED

In planning an evaluation, you should be clear on the questions you are asking. The following are important questions for you to consider:

How much growth in learning occurred because of the programme?

What aspects of the instructional programme contributed or took away from the results of the programme?

What aspects of the learning environment contributed or took away from the results of the programme?

How did the programme affect other people and the environment?

Is the programme's rationale valid and are the goals and objectives appropriate?

How do the results compare with those of similar or alternative programmes?

Once you have determined what important question or questions you wish to address, you will need to decide on an appropriate design.

DETERMINING THE EVALUATION DESIGN

The evaluation design establishes who and what will be evaluated and when data will be collected. Your choice of design is influenced by the questions you hope to answer. And as you will see, some designs yield data that are more objective, more apt to reflect real results, and more easily related to aspects of your programme.

Designs and Their Selection

Teacher evaluation frequently consists of a post-test design like that illustrated in Figure 3.

Figure 3
POST-TEST DESIGN FOR SUMMATIVE EVALUATION

In this design, you would test students at the completion of your programme. This kind of a test is known as a summative evaluation.

In a great many educational programmes, particularly those of several weeks or longer, it is usual for a number of shorter tests to be given during the programme (see Figure 4). These allow the teacher to check student progress along the way and make any necessary adjustments in the teaching process. This design, then, uses formative evaluation methods as well as summative.

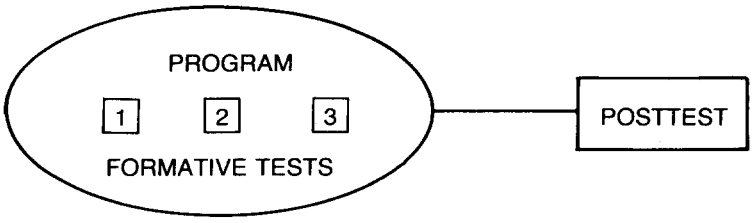


Figure 4
POST-TEST DESIGN FOR FORMATIVE AND SUMMATIVE EVALUATION

Teachers sometimes give a pre-test in order to find out where students are with respect to the programme's objectives. This also provides a benchmark for determining growth in learning or other effects when the results of the post-test are available (see Figure 5).

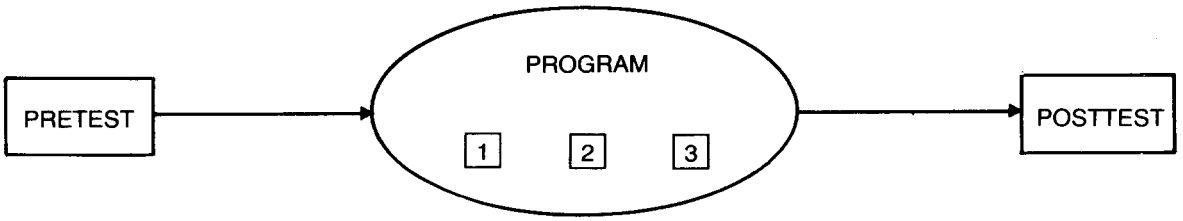


Figure 5
PRE-POST-TEST DESIGN FOR FORMATIVE AND SUMMATIVE EVALUATION

Other more complex designs involve comparing changes in students who participated in your programme with changes in students not in your programme (see Figure 6).

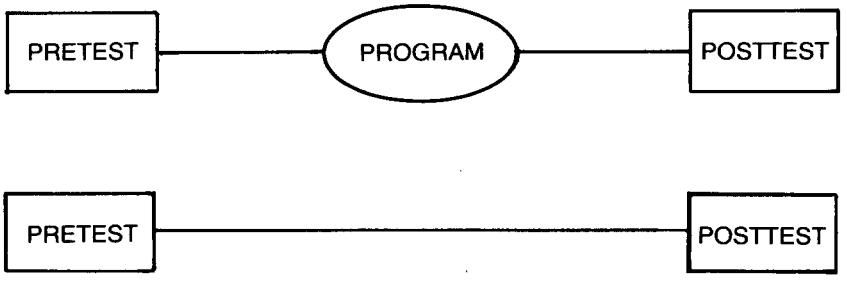


Figure 6
PRE-POST-TEST DESIGN WITH PROGRAMME AND NO-PROGRAMME GROUPS

These designs allow you to be more certain that the effects you observe are a result of your programme and not of something else.

Appendix A contains detailed information on choosing designs and other considerations for those who wish to explore this important evalution step further.

SELECTING AND/OR DEVELOPING THE EVALUATION INSTRUMENTS

Once the design is chosen you are ready to select and/or develop the instruments. These are the tools you will use to gather the data necessary to answer your evaluation questions. And since evaluation questions are more than likely concerned with how well you met your objectives, you will need to look at your objectives for direction in preparing instruments. And since unanticipated outcomes will also help you answer your questions, you will also need to search for these, too.

Above all, carefully think through the kind and amount of information you will need; it's usually difficult, if not impossible, to go back and collect missing pieces once the evaluation has been carried out.

General Considerations

Student/Teacher Planning

In so far as possible and with regard to the age and maturity of your students, consideration should be given to involving your students in planning the evaluation. Not only will you gain insight into their perceived needs and perceptions of the programme's goals and objectives, but you will be helping to establish a climate of trust and openness necessary for an effective programme. The importance of this with respect to the development of critical thinking was discussed in Chapter I.

Respect for Privacy and Other Student and Family Rights

You should be alert that your questions and test items do not exceed the bounds of respect for personal privacy. If you are not sure of particular items, it would be wise to seek an opinion from an administrator.

Effectiveness of Communication

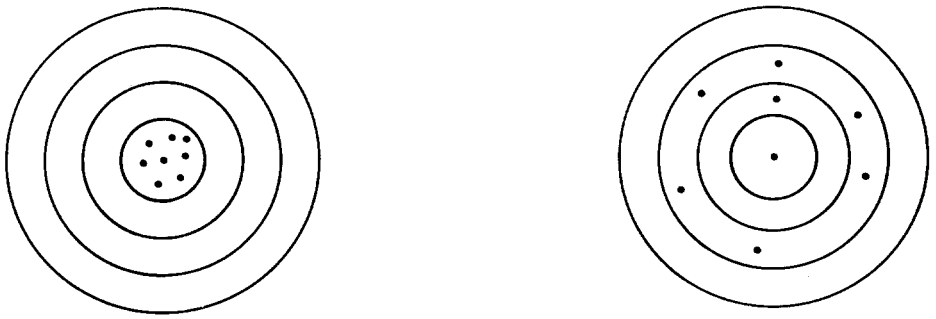
If your instruments are not understandable, then your results will be unreliable. Pilot testing (to be mentioned later) is of great help in determining how well the directions and questions are understood by those taking tests or answering questionnaires.

Validity and Reliability

Validity and reliability are two important characteristics of any evaluation instrument. You should have a grasp of the meaning and significance of these two concepts.

* Validity refers to the extent an evaluation instrument measures what it is designed to measure. For example, take the objective: 'The student will be able to identify three possible detrimental effects of acid rain'. If the instrument contains only items that test the student's understanding of what acid rain is but does not include items allowing students to identify detrimental effects, the instrument will not yield a valid measure of this objective. Another example: You may wish to measure the students' grasp of concepts, and your test is not valid because it only measures facts. An easy way to

picture validity is to think of a target. If your instrument has high validity, the test items will strike the bull's eye - what you want to measure (see Figure 7).



Valid Not Valid

Figure 7
VALIDITY

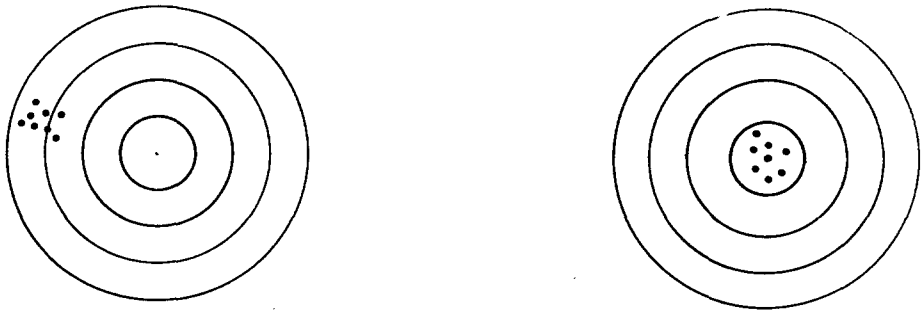
Content or descriptive validity is the type of validity most often examined. You can do this by having a group of your fellow teachers or other associates rate test items on, for example, a five-point scale rating the extent the items measure each objective. Items can then be reworded or substituted to increase their validity. Two different assessments of the same content can be correlated and a coefficient derived. When selecting instruments this information on validity is sometimes available. The following offers some guidance in assessing this information:

(Above .75) = high

(Above .50-.75) = acceptable

(Below .50) = uncertain.

* Reliability refers to the extent an evaluation instrument yields measures that are consistant each time it is administered to the same individuals. It is sometimes called reproducibility. If an instrument produces different results because of time or day or other conditions, it is unreliable. Note that an instrument can be reliable but not valid - the results can be consistent but still not measure what the instrument should be measuring (see Figure 8).



Reliable but not valid

Reliable and valid

Figure 8
REALIABILITY AND VALIDITY

One method of checking reliability is to administer the instrument twice with, perhaps, two weeks between each time it is used. Of course, since the time between the two administrations of the instrument is important, it should be reported; if given too soon, the second application of the instrument might yield higher scores. A correlation coefficient is calculated on many instruments to give a measure of reliability. As a general guide to assessing the reliability of a test when this information is given, the following may help:

(.80-.99) = high

(.50-.80) = questionable

(Below .50) = unacceptable.

A final word about validity and reliability: you will probably not have good information about the validity and reliability of the instruments you develop but do not let that deter you. Incorporate the suggestions presented here the best you can, seek the opinions of others, and use your practical sense and honest judgement. If you have access to a resource person with some background in measurement and statistical methods, you may wish to get some help. But do not let the absence of a knowledge in this area deter you.

Considerations in Selecting Instruments

Instruments should be selected on the basis of how well they are likely to measure your objectives and reveal unanticipated outcomes. Here are some key questions you should consider when selecting instruments:

Is the instrument appropriate for your class?

Does the instrument measure what it is supposed to measure (validity)?

Will the instrument give you consistent results (reliability)?

Is the instrument easy to administer and score?

Is the cost of the instrument and its use within your budget?

Is the instrument one which will allow you to compare your results with those of another similar programme or with national scores?

Considerations in Developing Instruments

Undoubtedly you will be developing most of your own instruments, although there are a few sources of instruments and test items available (see Chapter VII). Here are some basic considerations:

Give brief, clear directions.

Use an attractive, neat, legible format.

Be concerned with the tabulation of responses. Design the instrument so that responses may be countered and summarized easily.

Deal with only one concept, value, or skill with each item.

Avoid biased or leading questions. For example:

> Should we not teach about air pollution in our schools? A better way would be: rate the extent air pollution should be taught in our schools.

Check validity and reliability to the extent possible within your limitations of time, money and expertise.

Have your instruments reviewed by your colleagues and others, paying attention to the points outlined above.

Pilot test, evaluate and revise the instrument.

Pilot Testing

It is important to always test your instruments with a small group of students similar to those whom you actually plan to evaluate. This will give you an opportunity to check on aspects such as clarity, readability, length of time. You will also be able to check on ease of tabulation and data analysis.

It might be possible to test the instrument with another class or, if appropriate, only two or three individuals in a tutorial situation. The latter situation allows you the opportunity to observe a few students closely and question them carefully whenever you perceive a problem.

KINDS OF INSTRUMENTS RELATED TO OBJECTIVITIES

In this section you will be presented with the more traditional measures of objectives. In the next section, you will be exposed to measures used to explore more freely for unanticipated outcomes.

Growth in Student Learning

The following kinds of instruments and items are designed to measure student growth. Some are more suited to measure cognitive growth, whereas others are more appropriate for evaluation of attitudes and values and manipulative skills.

(a) Pencil and Paper Tests of Achievement

Pencil and paper tests are by and large the most popular evaluation instruments. They are particularly well suited to measuring knowledge, understanding and thinking skills. They are also used extensively to measure attitudes and values. But they are, of course, less suited to measure physical skills, except for skills such as mapping and illustrating.

It is important that you become aware of the limitations of pencil and paper tests as well as their advantages:

- A primary shortcoming is the 'guinea pig effect' or awareness of being tested. This may cause the respondent to answer to please or defy the teacher. This is of special concern when evaluating affective objectives and requires that pencil and paper measures be augmented by other more unobtrusive approaches which are not as likely to create attitudes.

- They may lack motivation and be viewed as having little relevance, thereby creating a less serious attitude.

- They generally favour verbal, reading and writing skills. Some objectives, of course, require these skills, but for others, students with a lack of verbal skills, may be prevented from giving an accurate picture of their achievement.

* Guidelines for development

1. Use a neat, easy-to-read design on the page.

2. Use clear, understandable directions for each response.

3. Provide specific spaces for responses (boxes, brackets, parenthesis, blanks) arranged on the sheet for easy tabulation, for example, along the edge of the sheet or on a special answer sheet.

4. Group questions requiring similar response formats.

5. Group items that are related such as those related to energy conservation.

6. Analyse each item to assess its validity, difficulty, and discrimination.

Validity - does the item measure what it is supposed to - has already been discussed under the section on validity and reliability.

Discrimination refers how well the item discriminates between high and low scoring students - those who have met the objectives and those who have not. First, divide the class into those who scored high on the test and those who scored low, For programmes with more than 50 students, select the highest and lowest scoring quarters. For programmes with less than 50 students, simply divide the group into halves based on high and low scores. Second, you would normally expect more high scoring students would do well on an

item than low scoring students. Those items that favour low scores may be measuring something different and may affect the test's validity and reliability. These should be identified and discarded or rewritten.

It is possible to quantify the amount of discrimination using the following formula:

$$\text{Discrimination} = \frac{\text{Number of high scorers who answered the item correctly} - \text{Number of low scorers who answered the item correctly}}{\frac{1}{2} \text{ the total number of high and low students}}$$

For example, using the previous example:

$$\text{Discrimination} = \frac{20 - 5}{25} = \frac{15}{25} = .60$$

The results may be interpreted using the following scale:

(-1.00 - .00) = nondiscriminating

(.01 - .20) = low discrimination

(.21 - .40) = medium discrimination

(.41 - 1.00) = high discrimination

Note that a perfect discrimination would be one. In other words, all the high scorers would get the answer right and all the low scorers would get it wrong:

$$\text{Perfect Discrimination} = \frac{25 - 0}{25} = 1.00$$

Difficulty may be checked using the following formula:

$$\text{Difficulty in percentage} = \frac{\text{(High) Number correct on the item} + \text{(Low) Number correct on the item}}{\text{Total number of high and low students}}$$

For example, if you have 100 students in the programme and of the top 25 only 20 got the question right and of the bottom 25 students only 5 students were correct, the difficulty level would be as follows:

$$\text{Difficulty in percentage} = \frac{20 + 5}{50} = .50$$

Interpret the results using the following scale:

(1.00 - .75) = easy item

(.74 - .26) = medium difficulty

(.25 - .00) = difficult item

* Sample items

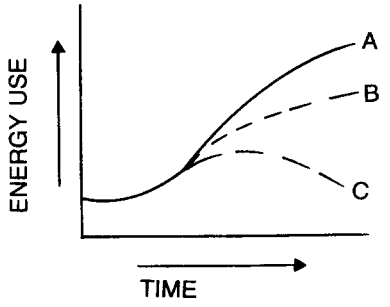
The following table presents sample items and additional considerations:

TABLE I: SAMPLE TEST ITEMS

Item type	Example	Advantages	Disadvantages and suggestions for improvement
True-false (or yes and no)	T F (circle 'T' for true and 'F' for false) In nature, the minerals in soil are picked up and used by plants. If nature is working properly, they will be reused	- East to construct and score - Useful when there are two alternatives	- Write true or false statements in a 50-50 proportion - Keep statements of uniform length - Avoid qualitative terms, such as all, only, always, etc. - Guessing may be a factor
Modifier true-false	T F (circle 'T' if the statement is true and 'F' if false. If false change the underlined word to make it true.) Transpiration is the loss of water from the surface of a lake	- Guessing is less a factor	- A completion type question is sometimes less ambiguous

Item type	Example	Advantages	Disadvantages and suggestions for improvement
Completion	(Write your answer in the blank at the left of the question.) (Pollutant) A contaminent in water which is detrimental to health is called a _____.	- Easy to construct - Good for short answers	- Difficult to limit answers to one correct response - Keep a list of acceptable responses
Multiple choice	(Circle the letter beside the correct answer.) In the picture, which of the areas is most ecologically stable? a. The area with trees b. The grassy area c. The bare ground d. The mud puddle e. The drainage ditch	- Promotes closure on questions - Gives options	- Provide five responses - Direct questions are better - Keep responses of equal length - Avoid responses that are negative - Best use where limited responses are desired - Construction is time consuming
Matching	(Match the items on the right with those on the left by placing the correct letter in the blank.) __1. Primary consumer a. Orchid b. Mouse c. Snake __2. Decomposer d. Fungus __3. Producer	- Easy to construct - Can deal with a number of related variables	- Have a different number of items in each column

Item type	Example	Advantages	Disadvantages and suggestions for improvement
Rating scale	(Mark your choice on the answer sheet.) How the land is used is up to each individual land owner 1. Strongly disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly agree or How do you feel about air pollution? 	- Provides opportunity to measure attitudes and values and degree of judgement	- Gives no information on judgement or reason - Directionality can be confusing
Rankings	(Rank the items below in your order of concern with 1 being your highest concern, 2 next highest, etc.) __ Planting trees on the school site __ Removing the incinerator __ Picking up the litter __ Creating a new playground	- Easy to develop - Provides a relative measure of interests, attitudes and values	- Limited by the number of similar topics

Item type	Example	Advantages	Disadvantages and suggestions for improvement
Semantic differential	(In a blank between each of the opposing terms place a check mark that indicates how you feel or what you believe) Food production (1) (2) (3) (4) (5) Important - Unimportant Successful - Unsuccessful Not a problem - A problem Simple - Complex	- Provides an opportunity to test both cognitive and affective objectives	- Terms must communicate
Modified essay	Explain the causes and effects of the three projected curves (dashed lines) showing the use of energy in a metropolitan area.  	- Requires less time to prepare - Evaluates a higher thinking skills and values and attitudes - Less subjective and open ended than essays	- Requires the identification of specific answers - Time consuming to analyse and tabulate

Item type	Example	Advantages	Disadvantages and suggestions for improvement
Essay	What do you see as the most serious environmental problem facing the world today and discuss the reasons for your concern?	- Requires less time to prepare - Evaluates higher thinking skills and attitudes and values	- Requires the identification of specific answers - Time consuming to analyse and tabulate - Correct the same essay question on each paper before proceeding to the next item

(b) Performance tests of achievement

Performance tests are tests in which students are asked to demonstrate an acquired manipulative skill, such as, calibrating or using an instrument. They allow you to approximate real situations more clearly then pencil and paper tests. Because they are a direct measure of the outcome, there is little inference required to determine if it has been achieved. However, performance tests may be time consuming to set up and administer.

* Guidelines for development

1. Determine the exact task to be performed and the setting and resources required.

2. Specify the criteria you will use to judge the degree of skill exhibited by the respondent, for example, length of time, use of equipment and quality of product.

3. Prepare clear directions.

* Sample items

1. On the topographic map draw the watershed surrounding the lake.

2. Using the equipment provided, determine the pH of the water to one-tenth of the pH scale.

(c) Questionnaires

Questionnaires are pencil and paper instruments used to measure interests, opinions and attitudes of individuals participating in or influenced by a programme. They may be administered like achievement tests, mailed to respondents for self-reporting, or used in interviews. Questionnaires are time consuming to

prepare and require careful pilot testing, but they can have great value over and above the data they yield. Used properly with the right questions, questionnaires can provide students with opportunities to give honest, meaningful input to the programme. They can represent a democratic approach to evaluation and help counter potential negative and oppressive effects that might be associated with the evaluation programme.

* Guidelines for development

1. Use a neat, clear, well-organized layout and format.

2. Use paper with soft colours - beige or light green; avoid harsh, bright colours.

3. Be sure of your audience.

4. Avoid trivial questions; ask for essential information.

5. Use simple, direct language in asking clear, specific questions.

6. Address only one topic at a time; avoid 'double-barrelled' questions.

7. Avoid leading questions.

8. Use open-ended questions when you wish to give the respondents freedom to express their own feelings, opinions, understandings, etc. Advantages include the likelihood of: (1) discovering unanticipated outcomes; (2) obtaining a wide range of responses; and (3) receiving candid answers.

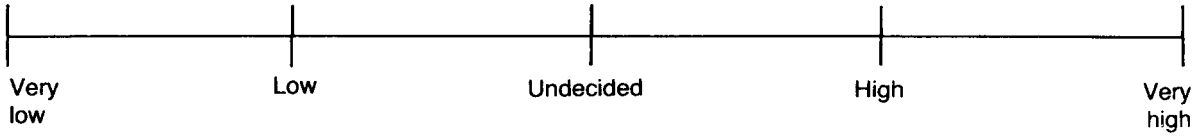
   Disadvantages include: (1) scoring difficult and time consuming; (2) responses may be dependent on verbal ability; and (3) the need to balance subjective responses with objective data.

9. Use focused items to force the respondents to make choices and give structured responses. Much of what has been mentioned about pencil and paper achievement tests applies here. An important additional point, however, involves the tendency of some respondents to choose certain responses. This is called response style. Response styles include the following:

   Social preference is a style in which individuals tend to choose what they think everyone else will and avoid questions about which society might have strong agreement or disagreement, for example, 'pollution is bad'.

   Passive agreement is a style in which some individuals tend to agree when they have a choice. Avoid questions like 'Do you agree that the cost of preventing acid rain should be borne primarily by those producing the sulphur compounds responsible for it?'. A multiple choice item might work better in this case.

   Rating scale bias is a style in which individuals tend to avoid or tend to choose extremes on a rating scale like that shown below.



This can be prevented to some extent by defining each rating in terms of a human response to the statement or question which is being rated. For example, take the question 'What priority should be given to cleaning up hazardous wastes?'. A very high rating choice might be defined as: government should give hazardous waste legislation first priority.

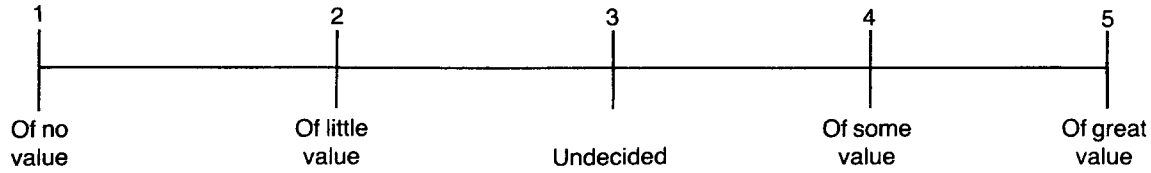
* Sample items

Open-ended item:

Is there an environmental problem or issue that concerns you? If so, why?

Focused item:

What is your overall feeling about the two-week unit on environmental issues of your community?



(d) Interviews

Personally conducted interviews are in reality live questionnaires and can have all of the same positive effects on the students' receptivity to the evaluation programme. They can be conducted in person or, in some cases, by telephone. Interviews may also be tape recorded for later analysis and checking.

Because you can probe and follow up your students' responses to an interview situation, it is possible for you to gain a deeper knowledge and understanding of mastery of concepts, acquisition of thinking skills, and development of attitudes and values. Students will feel more a part of the evaluation process. Interviews give you an opportunity to establish rapport and secure more frank and honest answers. You will be better able to detect bias and uncertainty. And you will be able to ask and explain more complex questions. Finally, this technique works particularly well with very young children and, in fact, is a necessary approach when children are too young to read. The major disadvantage of interviews is the amount of time they consume.

* Guidelines for development

1. Develop clear, easily understood questions.

2. Develop appropriate probes to use.

3. Develop a response sheet you can use to record data for ease in tabulation of responses.

4. Carry out several pilot tests of the interview.

* Sample items

In what ways do you think your style of living is compatible with a healthy environment? Explain.

Are there any ways you feel you would like to change your living patterns to help maintain an environment fit for life? Explain.

(e) Observational Instruments

Observing behaviour is a valuable evaluation technique. It is especially useful in helping you to assess a student's ability to apply understandings and skills, and it provides another way of determining the attitudes and values a student may hold. Observational information can supplement achievement data, both qualitatively and quantitatively. However, it is difficult to get valid and reliable data because there will always be a degree of inference involved as you translate behaviour into cognitive and affective outcomes.

Observation can be systematic or informal. Systematic observation involves using a detailed checklist of behaviours related to your behavioural objectives. It generally requires much practice.

Informal observation is open-ended requiring that you observe actions and events during class and make notes and summary statements immediately following. These may take the form of anecdotal records consisting of factual descriptions and separate interpretations.

Other means of recording class events are sometimes employed to supplement written records. Motion pictures, audio tape, and, more recently, video tape can be used. Still photos are another observational tool. Sometimes these are taken at regular intervals and judgements made on the frequency of behaviours observed.

* Guidelines for development

1. For systematic observation, identify specific behaviours from your objectives which lend themselves to observation.

2. Establish a rating scale for each behaviour. For example, the behaviour 'works independently in seeking solutions to the problem being studied' may be rated as good, adequate, poor and not observed. To further sharpen your observational skills, you should also define each rating.

3. Plan how the data will be recorded. For informal observations you may have special sheets, areas blocked out on a form, or notebooks.

4. Plan when and how often you will observe, for example, a five-minute period during the middle of the class or informally throughout the session.

5. Practice your observation skills one or more times.

* Sample items

An example of the use of observation and rating scales to assess attitudinal objectives comes from the Environmental Education for Guam Schools Project. Curiosity, inventiveness, critical thinking and persistence were rated on the following five-point scale: 1 - never; 2 - seldom; 3 - an average amount; 4 - more than average; 5 - to an outstanding degree.

Curiosity was defined in terms of four behavioural patterns: (a) using several senses to explore organisms; (b) asking questions; (c) observing organisms on first entering class; (d) showing interest in experimental outcomes.

Inventiveness was defined as: (a) using equipment in new ways; (b) suggesting new experiments; (c) describing novel conclusions from observations.

Critical thinking was defined as: (a) using evidence to justify conclusions; (b) pointing out weaknesses in others' reports; (c) changing ideas in response to evidence.

Persistence was defined as: (a) continuing investigations after the novelty has worn off; (b) repeating an experiment to get 'better' results; (c) completing an activity even though classmates finished earlier.[8]

(f) Artefacts

Students participating in environmental education frequently produce artefacts which can give evidence of the development of understandings, thinking skills, attitudes and values, and manipulative skills. These include investigative reports, letters, posters, bulletin board displays, articles, maps, photo studies, home-made environmental studies equipment, and environmental improvement projects such as school beautification, nature trails, and erosion controls.

* Guidelines for development

1. List artefacts likely to be produced by your programme.

2. Each artefact is the product of knowledge, understanding, thinking skills, attitudes, values, manipulative skills or some combination of these. Determine which of these the artefact represents.

3. Determine methods of analysis to assess kind and degree of knowledge, manipulative skills, etc., represented by each artefact, such as, rating scales.

* Sample items

1. TABLE II: EVALUATION OF A SLOPE MAP

Variables represented	Very poor	Poor	Average	Good	Very good
Ability to translate 3-dimensional forms into 2-dimensional representations					
Accuracy					
Etc.					

2. TABLE III: EVALUATION OF A RESEARCH PAPER SUPPORTING AN ISSUE

Variables represented	Very low	Somewhat low	Undecided	Somewhat high	Very high
Credibility of evidence					
Presentation of evidence					
Etc.					

(g) Unobtrusive Measures

Sometimes students will behave differently when they know they are being tested or observed. Unobtrusive measures do not intrude into the educational setting. Students are unaware that data for evaluation are being collected, thus, the chance that the results will be influenced by the collection procedures is eliminated.

These measures are particularly useful for indirectly measuring attitudes and values. For example, if one of your objectives is to increase student interest in environmental issues, you might check student use of library resources. Are students taking out books, magazines, and newspapers dealing with issues related to environmental concerns?

There are two kinds of unobtrusive measures: those that involve observations of activities in the normal school setting, such as the example described above, and those that involve observations of students in contrived situations. Contrived situations are those the teacher creates to test student reactions, for example, the planting of litter along a field trip route to see how many pieces will be picked up. Use of these kinds of tests must be well thought out to avoid negative feelings of being tricked or to avoid an unanticipated violation of ethical principles. They must also be well planned to avoid unexpected influences, since you have less control over the testing situation.

Unobtrusive measures can be a valuable part of your programme's evaluation. They are particularly useful in augmenting data from the measures described above and should be included in your evaluation design if possible.

* Guidelines for development

1. Examine a list of sources of unobtrusive data (such as that below) to determine which kinds may yield useful information.

2. Consider establishing contrived situations if appropriate.

3. Devise means of collecting and recording data from unobtrusive measures such as record sheets.

* Sample measures

Records of activities associated with the programme

1. Attendance rate, both required and optional.
2. Extra credit assignments completed.
3. Books and other materials taken from library or media centre.
4. Participation in extra-curricular activities.
5. Case histories.
6. Recognitions.
7. Leisure activities.
8. Courses selcted.
9. Tardiness.

Activities in contrived situations

1. Attendance at optional media programmes.
2. Volunteering to answer questionnaires of fictitious surveys.
3. Responses to bulletin boards, posters, or displays asking for voluntary participation such as entering a contest.
4. Comments heard in response to an environmental problem created on the school grounds such as the introduction of litter.

(h) Using Multiple Measures

If possible you should use several of the instruments and techniques described above to measure growth in student learning. The collection of data from parents, teachers, community members, and peers, as well as from the students in the programme, and the use of both direct measures and unobtrusive measures will give a much broader data base for making judgements. By using mutliple measures, greater confidence in your outcomes will be generated.

The Instructional Programme

The measures of growth in student learning described above, are the primary indicators of a programme's success or failure. And their use with some of the designs suggested earlier can allow you to equate student growth with both the entire programme and a particular instructional strategy such as a filmstrip, field trip, or simulation.

(a) Teaching Methods and Learning Activities

Measures of the effectiveness and efficiency of specific teaching methods and learning activities can be carried out during the programme as a part of a formative evaluation or conducted at the end as part of a summative evaluation. The principle variables with which you should be concerned are those that relate to your specific instructional activities - what you do and what your students do.

Variables to be measured that relate to instructional methods include: ease of use by the teacher, effect on students' interest and motivation, frequency of use, time required, impact on student learning, and cost.

The data from variables related to instructional methods, along with measures of student growth, can help measure both the efficiency and effectiveness of specific methods.

* Guidelines for development

If you want to measure the efficiency and effectiveness of specific method:

1. Identify the method you wish to evaluate.

2. Select a suitable design (see the section on designs and Appendix A). For example, will you use a pre- and post-design, a design in which you measure effectiveness throughout the teaching, or a post-test-only design? Your design will help you decide whether you will collect and use the data - during the programme (formative evaluation) or at the end (summative evaluation).

3. Decide what important variables are related to the method you wish to evaluate.

4. Develop measurement instruments and items.

5. Schedule times of actual data gathering.

* Sample items

Tests of achievement - pencil and paper and performance

These tests may be given immediately after a particular method is used. They measure how well the method achieves your objectives.

Questionnaires and interviews

These evaluative techniques allow you to explore student opinions regarding the instructional programme and the methods used as well as to probe for cognitive and affective outcomes, either specified in the objectives or unanticipated. For example, you might ask the questions: What did you like best? Why? What didn't you like? Why? What did you learn?

Observational instruments

Checklists, logs, case-study records, and other kinds of data-recording instruments for observations allow you to describe the use of instructional activities and observable responses of students during the actual teaching. For example, you might note spontaneous negative and positive comments, length of time to become involved, and questions asked.

Artefacts

The quantity and quality of artefacts produced by the method or activity allow you to infer the interest and motivation generated by the activity, the understanding and thinking required, and the skills developed. For example, plans for beautification of the school can demonstrate a knowledge of materials, an understanding of design principles, and artistic skills.

Unobtrusive measures

Examples of these measures include: noting how many students exhibit an interest in continuing the activity on their own or after hours, discussion with peers, and participation in optional follow-up activities.

(b) The Learning Environment

The learning environment consists of the teacher, students, and the setting of the classroom, school and community. All of these make up the context in which the instructional programme is introduced.

Those variables to be assessed that relate to the teacher include: self-concept, training and experience, philosophy of teaching, teaching style, theory of learning, skills in interpersonal relations, attitudes towards the students, belief in the programme, including its goals and objectives, and concern for the environment.

Variables that relate to the students include: ability, age, sex, prior achievement, attitudes towards the teacher, school and learning, peer relationships, socio-economic background.

Variables that relate to the classroom environment include: creature comforts (lighting, noise level, temperature, humidity), kinds and location of furniture, equipment and media materials, kinds and availability of supplies, space and ease of movement, opportunities for independent study, small group interaction and whole-class activities (seminars, lectures, demonstrations).

Variables that relate to the school include: general support for the programmes, attitudes and policies towards valuing activities, field experiences, school organization, per pupil expenditure, size of the school, student/teacher ratio, opportunities for nearby outdoor studies, and general maintenance and care of the school environment.

Variables that relate to the community include: encouragement and support for action projects, numbers of community volunteers, opportunities for environmental studies and field trips, and co-operation of public and private groups.

Data about these and other variables of the learning environment, can give you insights as to why certain instructional methods worked or did not work, why student growth occurred or did not occur, and why the programme had a positive impact on the environment or did not. You can record your own observations and perceptions on self-evaluation instruments as well as collect data from other school personnel, community people, and your students.

* Guidelines for development

1. Identify those variables you feel will have the most influence in your programme.

2. Select a suitable design (see the section on designs and Appendix A), such as a post-test-only design.

3. Develop instruments and items.

4. Schedule times of actual data gathering.

* Sample items

Questionnaires and interviews

These techniques allow you to develop and use questions to ask yourself and others about the influence of variables you have selected. For example, what policies and practices does the school have for its own waste disposal or energy conservation?

Observational instruments

Through the use of these measuring devices you can rate the existence, quality, and amount of many variables, such as, media resources, space and equipment.

Unobtrusive measures

Unobtrusive measures allow you to gather information about important variables related to teachers, students in the programme, and participating community citizens. For example, attendance at meetings, correspondence, memberships and voluntary assistance all give indications of the climate in which your programme operates.

(c) External Effects

Ultimately your environmental education programme, if successful, should have positive effects on other people, encouraging their support and environmental responsibility, and it should directly affect the environment in positive ways. For example, is your community cleaner, safer, more aesthetic because of your programme? Measures of these and other variables should not be neglected because data of this kind, if believable, can bring a great deal of satisfaction to you and your students as well as increase support of the programme by the school and community.

* Guidelines for development

1. Identify the kinds of evidence that you will look for to determine the extent your objectives are achieved. For example, is the park cleaner because of your programme's efforts? Is your water monitoring programme receiving positive comments? Are there letters appearing in the newspaper supporting your activities?

2. Select a suitable design (see the section on designs and Appendix A), such as, a pre-post observation design.

3. Develop instruments and items.

4. Schedule data-gathering activities.

* Sample items

Effects on people

Questionnaires and interviews provide an opportunity to question citizens about their opinions and activities. You may also have the opportunity to observe them directly, for example, in volunteer community projects. Artefacts, such as letters and photographs, are also useful. Many of these may be unobtrusive, for example, records and attendance at meetings.

Effects on the environment

In a sense, environmental effects are artefacts. Checklists, descriptions and before-and-after photographs are a sampling of observational records you may use.

INSTRUMENTS TO DETECT UNANTICIPATED OUTCOMES

Even if it were possible (and it isn't), it would be unwise to anticipate and develop objectives for all the outcomes of your programme; it would be too time consuming, overwhelming, frustrating, and constraining. A more reasonable stance advocated here in this guide is to develop only a few measurable outcomes, then use a variety of measures to gather data about other outcomes that you cannot anticipate. It is quite possible that unanticipated outcomes will be more valuable in helping you judge the worth of your programme than those you spent hours trying to anticipate.

Kinds of Measures

Those measuring devices and techniques which allow you to probe and explore, to search out, are the ones to use in seeking the unknown - what you do not anticipate. From them you can gain both qualitative and quantitative information.

Questionnaires and interviews with open-ended items are extremely useful. Some examples of questions to ask: What was the most important thing you gained or learned? What do you wish had been done differently? What gave you your greatest satisfaction? Disappointment? How do your friends feel about the project? In what ways do you think the community benefited?

Carefully developed observational instruments with open-ended items can help you to become more alert to seemingly insignificant changes or the sometimes overlooked obvious. For example, in what ways are parents supporting the programme? What are other students saying about the programme's activities? Many of these approaches may use unobtrusive techniques involving both the study of records and artefacts and contrived situations such as the completion of an application to become a member of a fictitious organization.

The case-study approach is also a useful means of detecting unanticipated outcomes. In this approach, you would use data obtained from the techniques described above and describe in detail what has happened to your students, others in the community, and the environment as a result of your programme. Frequently, as you pour over data and attempt to set it down in words, you discover new things about it and gain new insights that you did not anticipate.

Action research approaches in which you examine your programme and its effects also provide a means of detecting unanticipated outcomes. These include in-depth interviews with your students; written, photographic audio, and videotaped records of your classes; involvement in discovering with your students; and using invited observers.

Measures to Judge the Appropriateness of the Programme's Rationale, Goals and Objectives

All of the measures discussed will yield data which should be used, periodically, to assess the need for the programme and the appropriateness of its goals and objectives. In addition, you should also consider asking others for their opinions - people whom you respect and who will give you candid views.

Your search for unanticipated outcomes may also reveal unanticipated needs you did not realize you were serving and, thus, give you additional rationale for the programme and new goals and objectives. Your data may also indicate that you cannot realistically accomplish some goals and objectives, in which case they should be discarded or modified.

Preliminary Plans for Analysis of Data

When you develop instruments, test items and questions, you should also plan in advance the kind of analysis you will be using. This will allow you to set up the instrument and data sheets for easy tabulation and treatment of the data. For example, you might develop categories to tabulate the responses to open-ended questions on a questionnaire. For the question, 'How did you feel about the activity?', you could have boxes for negative and positive responses on the tabulation sheet. These would help you to determine more quickly frequencies and per cent distributions.

Pilot testing the development of instruments often provides an opportunity to also do a dry-run of analysis procedures. Since tabulation and analysis can be time consuming, anything you can do to anticipate what will be required for this important step will more than pay for itself.

CHAPTER III: HOW CAN I CONDUCT MY EVALUATION? (STEP THREE)

There are two major steps in conducting your evaluation: collecting the data and analysing the data. Once these have been completed, you will be ready for the final step, using the results. This chapter clearly presents each procedural step, including the options you have for detours and side trips as you collect and analyse your information on the programme and its results.

COLLECTING THE DATA

Since this guide assumes that you, the teacher, have a central role in the evaluation, you can have direct control over the consistency of your approaches for collection of data. When several people are involved, this can be a problem unless, of course, they are all your clones working in identical situations. Since this will unlikely be the case, be advised to be consistent in administering pre- and post-tests and using such measuring techniques as checklists and interviews. Merely a word might create a different response and affect the validity and reliability of your results. If you do use assistants, give each the same, specific directions.

You should also attempt to anticipate potential problems in the data collection process. The following considerations provide a kind of checklist to carry out the procedure smoothly:

General Considerations

1. Establish a time schedule for collection of data. Take into account influences such as the effect of vacations (it is usually wise to avoid the day before or the day after), exam weeks, and special events in the school calendar.

2. Obtain necessary approval of all arrangements from administrators and other personnel - secretaries, other faculty, and volunteers. This will involve communicating purposes and expectations and working out schedules.

3. If you are using assistants, select those who will tend to be objective. Give each an orientation, discussing the need for consistency, orderliness, awareness of personal bias and subjective influences, the influence of personality, and the need to monitor the process closely. Practice sessions should be encouraged.

4. Monitor the data-collection process closely. Be sure that plans are carried out as you planned with careful attention given to having enough instruments, providing directions, allowing the proper amount of time, noting incidents that might influence the results, and recording data accurately and correctly.

5. In the case of absentees, it may be possible to follow up in cases where questionnaires in surveys are used, but care will need to be used to avoid the influence of additional time and other variables.

Considerations Before, During and After

The following provide a brief checklist for the administration of each instrument:

Before the collection of data:

1. Study and review the instrument, data sheets and directions.

2. Practice using the instrument and rehearse the process.

3. Check to see that necessary materials are available and ready.

4. Check the setting - where the evaluation is to take place - to remove clues, distractions and other sources of influence.

During the collection of data:

1. Be alert to possible disruptions and the comfort of those being evaluated.

2. Keep a neutral personality; be interested in the process but take care that you do not intrude into the setting and bias the attitudes of the respondents.

3. If you need to give directions, speak clearly, slowly and loud enough for all to hear. Follow directions exactly.

4. Be sure each respondent has all materials.

5. Keep track of time requirements.

After the collection of data:

1. Collect and count all instruments and data sheets.

2. Briefly check all answer sheets and other data sheets for completeness.

3. Record any special incidents or events that could have influenced the results.

4. Organize the data for analysis.

ANALYSING THE DATA

At this point you may now be perched on a mound of data seriously considering abandoning ship before you sink out of sight in a sea of paper. The following pages are designed to help you easily and effectively deal with the plethora of information you may have accumulated.

Analysis involves the organization, summarization and interpretation of your data using numbers, mathematical operations and word descriptions. The mathematical operations of analysis are called statistics. There are two types of statistics: descriptive and inferential. Descriptive statistics are numbers that describe data.

For example, they include how often something occurred (frequency) and what proportion something is of a whole (percentage). Inferential statistics are numbers that help you to test your guesses and make inferences - conclusions from the data - about the effectiveness of your programme.

Through analysis you can:

determine achievement of your goals and objectives;

determine unanticipated outcomes and their relative values;

determine the worth of your goals and objectives.

Together, these roles of analysis enable you to judge the worth of your programme. Although the techniques presented in this chapter will help you make these determinations with increased objectivity, you will not be able to eliminate subjectivity altogether. Indeed, you should not, since intuition has, in more than one instance, kept programmes alive which eventually were proved effective in unforeseen ways. And, in fact, if you have followed these guidelines this far in evaluating your programme, you will have accumulated a fair amount of subjective data which may reinforce your own intuitions.

How to Summarize your Data (descriptive statistics)

In this section you will be introduced to some ways of reducing your piles of questionnaires, tests and checklists into neat summaries and numerical descriptions for easy interpretation.

(a) Describing Data Graphically

Graphs are visual summaries of information that make data more understandable and interpretable. Several such schemes are shown below:

1. Histogram (bar graph)

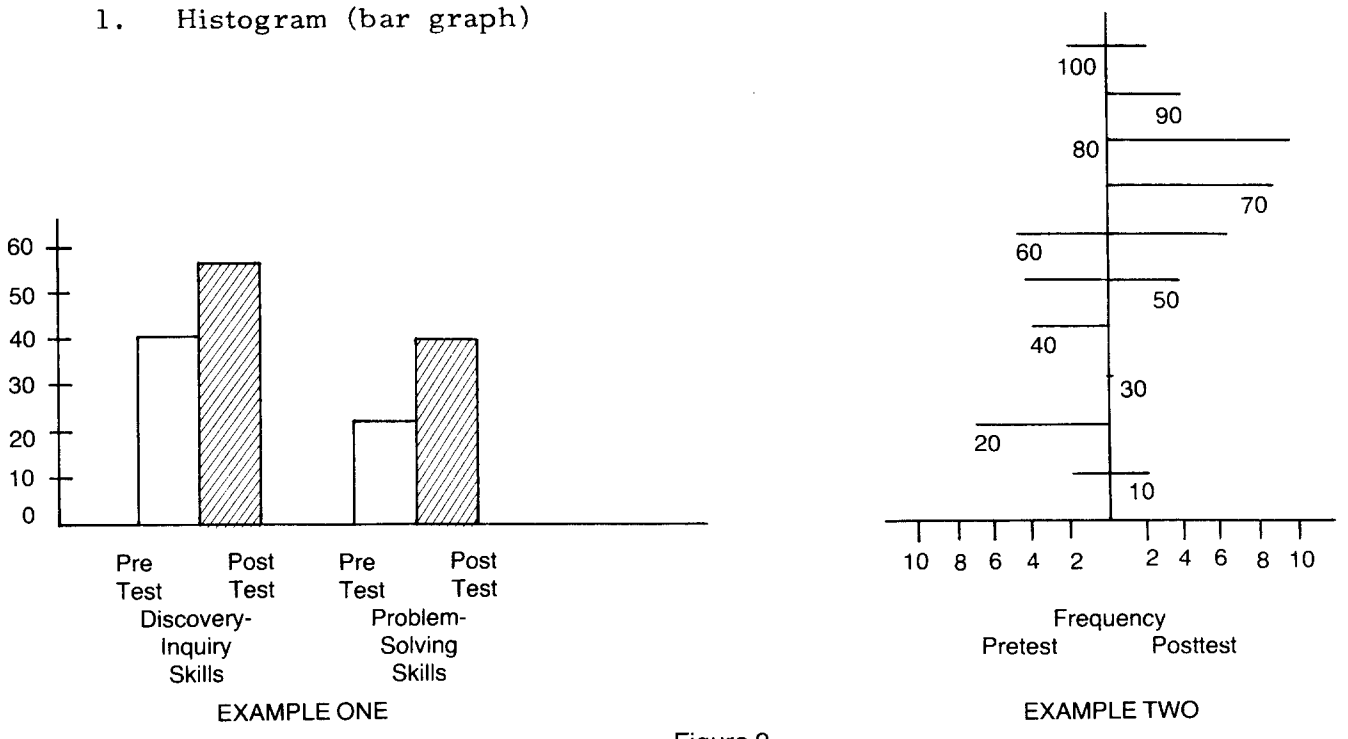


Figure 9
HISTOGRAMS

2. Frequency polygon

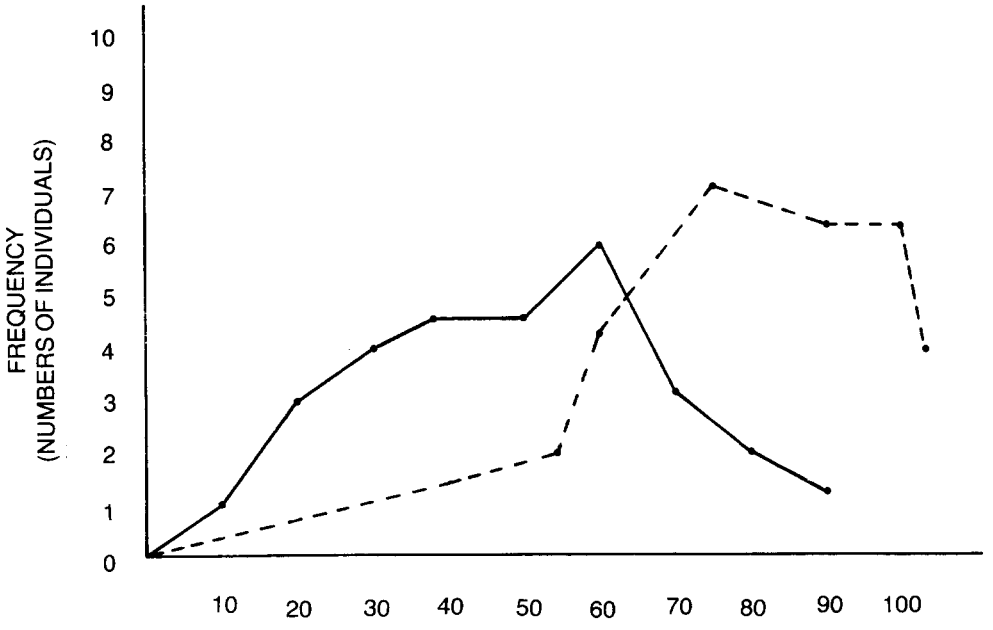
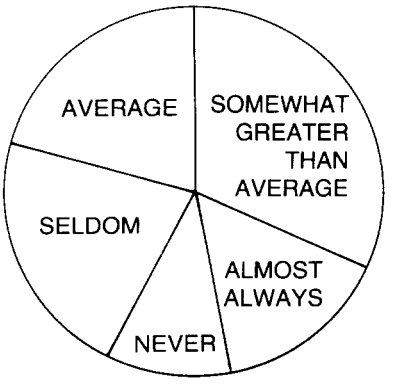


Figure 10
FREQUENCY POLYGON

3. Pie graph



PERCENTAGE OF STUDENTS DEMONSTRATING ATTITUDES
OF ENVIRONMENTAL RESPONSIBILITY

Figure 11
PIE GRAPH

The pie graph is useful in depicting proportions. To construct, multiply the percentage of each category times 360° and lay out the circle with a protractor.

(b) Finding Frequencies

A frequency is the number of times that something occurs in a given time or in a number of possibilities. For example, on a test of critical thinking skills, five students scored 10 correct out of a possible 50, seven students got a score of 20, and 15 students received a score of forty. Frequencies are used as a way of quantifying data into categories - a way of beginning to reduce the data into more useful forms. If, for example, you sent home a questionnaire asking parents to describe ways your students are demonstrating greater responsibility towards their environment, you would first need to categorize the responses. Categories might include: comments more on news items, does not waste as much electricity, and so on. As you review the answers, place a check in each category. The total number in each is the frequency. This kind of analysis is an example of content analysis.

(c) Findings Percentages

A percentage is a fraction or ratio of 100. The denominator would then be 100, for example, $\frac{15}{100}$. The 15/100 can be converted into a decimal by dividing 15 by 100 =

$$
\begin{array}{r}
.15 \\
100\overline{)15.00} \\
\underline{10.00} \\
5.00 \\
5.00
\end{array}
\quad = 15\%
$$

Frequencies are often converted to percentages to see more clearly the proportion of whatever is being measured to the whole. For example, you may have your class rate a film on ecology. The table below illustrates how frequencies and percentages are used to analyse the ratings:

TABLE IV: STUDENT RATINGS OF A FILM (24 STUDENTS)

Ratings	Frequencies	Per cent
Very good	12	50
Good	6	25
Undecided	2	8
Poor	4	17
Very poor	0	0
Total	24	100

(d) Finding a Middle Value

When you are faced with a range of scores, as you might be after tallying a test, one naturally asks, how did the group as a whole score? A common approach is to find a middle value, often called a measure of central tendency. Three commonly used measures of central tendency are described below:

* The arithmetic average or mean

The average or mean is the most common indicator of central tendency. It is simply calculated by adding the scores in a distribution and dividing by the number of scores. Assume, for example, that you have given a test of ten items to your ten students and that you get the following results:

RESULTS OF TEST

N = 10 (ten students)

	92		
	87		
	85		
	84		
	84		
	84		
	80		
	80		
	14		
	5		69.5 = MEAN
Total	695	10 ⟌	6950

* The midpoint or median

The median is another common index of central tendency. It is the point between the two halves of a set of scores. Counting up five or down five in the set of ten test scores above, the median would be eighty-four.

* The value occurring most often or the mode

The mode is the most frequently occurring score. In the distribution above it would be eighty-four. In this case it is the same as the median.

* Determining which to use

Generally the mean is preferred. It is the only one of the measures that uses all the data in a distribution. However, in the example given, the two very low scores affected the mean substantially giving a distorted, much lower score than the students generally achieved. In this case, the median and the mode both gave better measures of central tendency. Because of the effect of how the scores are spread out - the variation of the scores - you may wish to compute all three measures before deciding which to use.

(e) Finding an Index of Variation or Spread

As you can see from the test scores used in the previous example, the variation or spread of the distribution of scores can affect the measure of central tendency. Therefore, this measure alone cannot give a complete description of your data; you should also consider the amount of spread or variability. Two common measures of variability, sometimes called dispersion, are described below:

* The range

The range is simply the difference between the top and bottom scores. In the previous example, the range is calculated by subtracting 5 from 92. The difference, 87, is the range. You can see that the range is only a rough measure of variability; the next method yields better indices of variability.

* The standard deviation

This statistic can tell you more about variation by giving a figure that tells how closely the scores cluster around the mean, or average, score. The closer the scores cluster around the mean, the smaller the standard deviation.

This index is probably the most widely used measure of variability. It is particularly useful when comparing two groups. If, for example, group A has a standard deviation of 10 and B has a standard deviation of 6, you would know that the more students in group B have scores closer to the mean.

If you graphed the scores from the test in the previous example, they would look like the following: (see Figure 12)

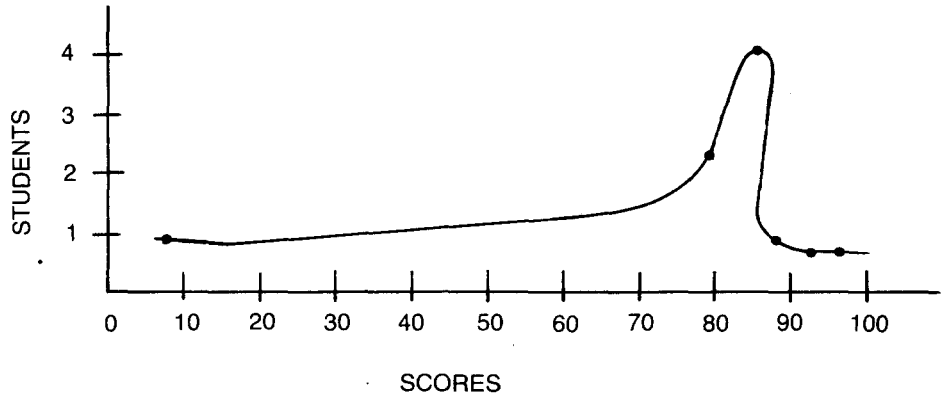


Figure 12
GRAPH OF TEST SCORES

To compute the standard deviation use the formula: (see Figure 12)

$$\text{standard deviation} = \sqrt{\frac{\text{sum of squared deviation scores}}{\text{Number of scores}}}$$

The process is as follows:

Step 1: Compute the mean, which is 69.5.

Step 2: Subtract the mean from each score to get a series of numbers that show how much each deviates from the mean.

92 - 69.5 = 22.5
87 - 69.5 = 17.5
85 - 69.5 = 15.5
84 - 69.5 = 14.5
84 - 69.5 = 14.5
84 - 69.5 - 14.5
80 - 69.5 = 10.5
80 - 69.5 - 10.5
14 - 69.5 = -55.5
5 - 69.5 = -64.5

Step 3: If you added the deviations above, they would add up to zero because two are minus numbers. So to get the average of the scores you must first square each. This will get rid of the minuses or negative numbers (a negative number times a negative number yields a positive number, e.g. -5 x -5 = 25). Then divide by the number of scores to get an average deviation. See below:

$$
\begin{array}{rrrr}
22.5 & \times & 22.5 & = & 506.25 \\
17.5 & \times & 17.5 & = & 306.25 \\
15.5 & \times & 15.5 & = & 240.25 \\
14.5 & \times & 14.5 & = & 210.25 \\
14.5 & \times & 14.5 & = & 210.25 \\
14.5 & \times & 14.5 & = & 210.25 \\
10.5 & \times & 10.5 & = & 110.25 \\
10.5 & \times & 10.5 & = & 110.25 \\
-53.5 & \times & -55.5 & = & 3,080.25 \\
-64.5 & \times & -64.5 & = & 4,160.25 \\
 & & & & \overline{9,144.50}
\end{array}
$$

$$\frac{9,144.5}{10} = 914.45$$

Step 4: Now, to get everything back to the size of your scores - the numbers you were working with originally - you must take the square root of the average of the summed squared deviation.

$$\sqrt{914.45} = 30.24$$

One standard deviation unit = 30.24

Now you can see by the formula

$$\text{S.D.} = \sqrt{\frac{\text{sum of squared deviation scores}}{\text{number of scores}}}$$

that the greater the distance the scores are from the mean, the more spread out or dispersed they are, and the larger the standard deviation will be.

Mathematicians have figured out that if a great many people took a test the normal distribution of scores, by chance, would look like the following: (see Figure 13).

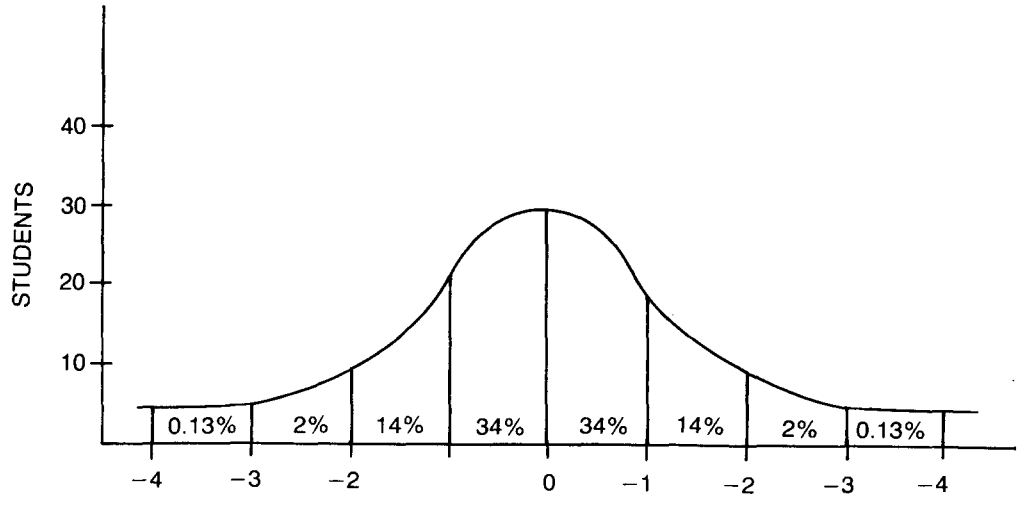


Figure 13

THE NORMAL CURVE

They have further found that one standard deviation unit is equal to about 34 per cent of the scores, so with a mean of 50 and a standard deviation unit of 15, approximately 68 per cent of the students will score between 65 and 35. About 96 per cent of the scores will occur within two standard deviation units on each side of the mean, and almost all lie between three units.

Now, returning to your original test, you know that with a standard deviation unit of 30.24 and a mean of 69.50 that 68 per cent of the scores lie between 99.75 and 39.25. So knowing the standard deviation tells you that there are a few low scores below 39.25 and that the curve is skewed to the right or upper levels.

Analysing Data for Student Growth

In this section you will be introduced to methods of using descriptive statistics to determine student growth. As a teacher you are eager to find out how your students do both individually and as a group on tests and other measures. Not only do you want them to succeed, but you know their performance is probably, to a large extent, a reflection of your teaching. How can you judge the achievement of each individual and the group as a whole? What reasonable standards can you use to say that this student's performance was excellent and is worth an 'A' or a 'B', or the group was not exceptional but met an acceptable level?

In this section you will be shown how statistical summaries and descriptions can be used to help you judge the performance of your students against some reasonable criteria. First, your behavioural objectives can provide one such set of criteria. They can establish a basis for grading the performance of each student or the group as a whole. Second, you can also use summaries of the scores of your class as a basis for judging the achievement of individual students or comparing your class with other groups. Finally, among your data there may be unanticipated outcomes, such as, the development of a concept or an attitude, you did not expect. You will need to judge the value of these, and this is also discussed.

(a) Evaluating Student Growth Using Objectives as a Basis for Comparison

One set of standards against which to judge the performance of your students is your behavioural objectives. They provide a basis for determining student growth at either the individual level or group level.

(b) Determining Individual Achievement of Objectives

This analysis is especially appropriate when diagnosing individual needs and prescribing individual instruction.

A typical objective would be as follows:

The student will be able to demonstrate mastery of the concept of ecological stability by responding correctly to 60 per cent of the items on a written test.

The analysis would only require the tabulation of scores and computing the percentage of correct scores. If the test had 30 items and the student got 24 correct the fraction would be $\frac{24}{30}$ or 80 per cent - well above your standard of 60 per cent.

(c) Determining Group Achievement of Objectives

The analysis of how your class as a whole achieves the objectives is a better measure for evaluating your programme.

A typical objective would be as follows:

> Of your class, 75 per cent will be able to demonstrate mastery of the concept of ecological stability by responding correctly to 60 per cent of the items on a written test.

This analysis requires, first, finding out how many students got 60 per cent of the items correct. Let us say 15 of 25 students got over 60 per cent correct. The percentage of those getting 60 per cent correct is $\frac{15}{25}$ or 60 per cent - not high enough to meet your objective. In this instance, you would need to investigate the reasons further, for example, you might do an item analysis to check the test or interview low scorers to find out why they did not reach your expectations.

(d) Evaluating Student Growth Using Measures of Class Achievement as a Basis for Comparison

Another common method of judging the performance of your individual students or the class as a whole is to compare them to an expected mean or other measure. For example, you might say:

The average score or mean of the group will be considered an acceptable level of achievement and anything below the mean by a certain number of points will be considered failing. For example, if you give a 100-item test to 25 students and the mean is 80, you might pass all students who receive scores below 80 that are within 25 per cent of the mean. So 25 per cent x 80 = 20 and 80 - 20 = 60. Hence, any score below 60 would be too low to be acceptable. Although setting these kinds of standards appears to be quite arbitrary, in reality this method usually reflects the teacher's observation of many factors, including ability levels, amount of time spent on the topic, and past performance.

Sometimes, teachers are tempted to use standard deviation units as a means of establishing grades, for example, giving a grade of C if it falls within one standard deviation unit of the mean. B's and A's might be given for those within the second and third standard deviation units above the mean. This practice is discouraged; the normal curve is a very arbitrary standard that does not take into account other factors which you might wish to consider such as encouraging future performance or rewarding improvement. Furthermore, the normal curve is only an idealized mathematical distribution which is only approached as one gets more and more scores. For small classes, the normal curve will normally not exist.

(e) Comparing the Achievement of Groups

The measures of mean, median, mode, range and standard deviation are measures which allow you to compare one group with another. For example, pre-test means and standard deviations of a group can be compared with post-test means and standard deviations. Likewise you can compare the means and standard deviations of programme and non-programme groups.

Ideally you would hope that the curve suggested by the mean and standard deviations for a post-test group would be higher on a distribution of scores than that of the pre-test. See Figure 14.

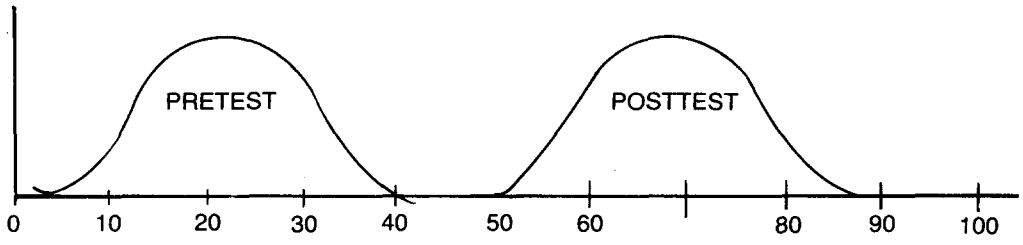


Figure 14

SEPARATED PRE-POST-TEST CURVES

Clearly in the above example there is a real difference. Usually, however, the curves are overlapping and determining whether or not there is a real change is less certain. See Figure 15.

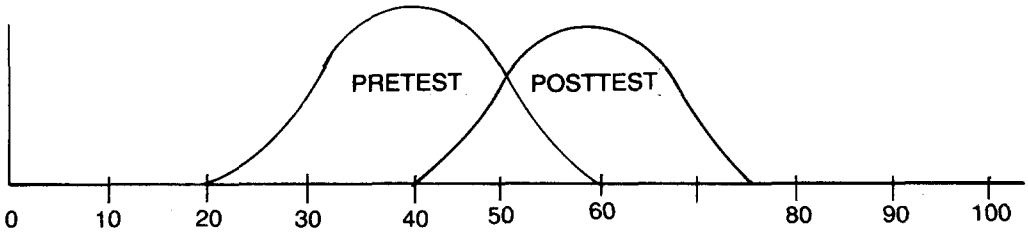


Figure 15

OVERLAPPING PRE-POST-TEST CURVES

But if you should have small standard deviations, then you will know that the scores cluster more closely around the means and there may be less overlap, even though the means may be separated by only 20 points or so. See Figure 16.

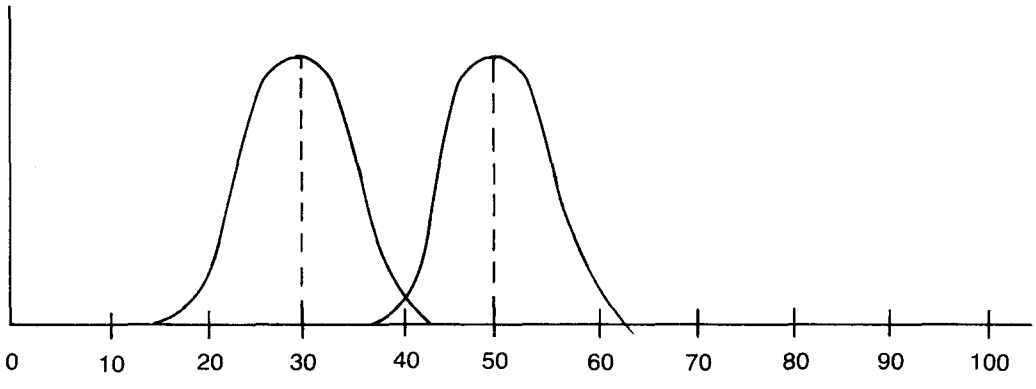


Figure 16

OVERLAPPING CURVES WITH SMALL STANDARD DEVIATIONS

To determine the probability that real distances (differences) exist between two or more groups, you will need to use inferential statistics. In other words, inferential statistics can help you decide whether your results are a product of chance or a result of some other factor, hopefully your programme.

Since the typical classroom teacher, for which this guide is intended, has less need for inferential statistics and their use requires more training and expertise than the use of descriptive statistics, you are referred to any good statistical text or, better still, to someone well versed in inferential statistical methods.

(f) Judging the Value of Unanticipated Outcomes in Student Growth

Throughout this guide you have been encouraged repeatedly to seek evidence of unanticipated outcomes such as new concepts, attitudes or skills you did not expect. This kind of evidence is likely to show up in essay items on pencil-and-paper tests, during observations, or in responses to open-ended questions in questionnaires, interviews and self-evaluation instruments.

Suppose, for example, that you received a series of responses you did not expect from the question,'What do you feel is the most important thing you learned?'. In this case you would first do a content analysis, setting up categories for different responses and recording the frequencies of responses in each category. For instance, 10 students of a class of 30 might indicate that they felt the most important thing they learned was that they could really make a difference in changing the attitudes of their parents towards an environmental issue. In this case you would also want to calculate the percentage, 33 per cent, so that you could easily compare this attitude change with other changes.

Analysing Data for Programme Improvement

As discussed previously, one of the major reasons for evaluating your programme is to find out how successful your programme is and what might be done to improve it. Designs and instruments have already been described to help you attribute results to your programme as a whole or to specific teaching methods and aspects of the learning environment. At this point it is assumed that you have a variety of data at hand and are ready to analyse it in terms of what worked best and what did not or what had the greatest or least effect on your outcomes. There are several analyses which you might be in a position to make.

(a) Analysing the Results of Post-tests to Determine Aspects of the Programme Responsible for Outcomes

As described in the last chapter there are a number of methods that yield data useful in determining what aspects of the programme are responsible for changes. These instruments include: (1) tests given immediately following the application of a particular teaching method, such as a simulation or film; (2) observations, while the programme is in operation; (3) artefacts created during instruction; and (4) questionnaires or interviews through which students report their opinions about what they felt was the most effective teaching method or factor in the learning environment.

As a general rule you will want to categorize your data, determine frequencies, and calculate percentages. In the case of tests, you will also want to use some of the other descriptive statistics.

(b) Comparing the Results of Pre-tests and Post-tests to Determine Gains Attributable to the Total Programme

Hopefully, through pencil and paper and performance tests of achievement and unobtrusive tests, you will see post-test scores higher than pre-tests as a result of your programme. As previously discussed under the section, Comparing the Achievement of Groups, a number of descriptive statistics may be used. Should you wish to go beyond descriptive statistics and use more powerful statistical analyses to determine the statistical significance of your results, you are encouraged to seek outside help.

There is, however, one important test of significance which does not require a test of statistical significance. This is a test of the practical significance of your results. Such a test is really a judgement as to whether or not the differences are large enough to be worth while. Is a 10 point gain, for example, large enough in your mind to be of value? This judgement must come from your own logic and reasoning or that of others; statistical procedures cannot help you. Inferential statistics can only tell you the likelihood that a real difference exists between groups and the probability that the scores occurred by chance or because of your programme.

(c) Comparing Evaluation Results of Programme and Non-programme Groups

A more persuasive relationship between your programme and measured outcomes can be obtained when you compare the post-tests of groups having the programme with those not receiving it. To determine the influence of a particular teaching method or aspect of the learning environment, you can compare the post-test of a group exposed to the particular variable under investigation with the post-test of a comparable group exposed to a different variable, or that of a comparable group not involved in the programme. So, for example, imagine that you have the scores from a test on water pollution that was given to three groups in your class: a small group that saw a film on water pollution, a group that read a chapter in a textbook and a group that received no special instruction. Of course you will need some assurance, such as a pre-test, that the groups did not differ significantly before the programme. By comparing the means and standard deviations you can make some judgements on the effects of each method. But inferential statistics may be of value in giving you greater confidence in the differences.

A useful statistical test for determining if certain results are related to the programme is the chi-square test. Suppose, for example, that you want to find out if your programme has created a concern for air pollution. You ask the programme and non-programme students to respond yes or no to the question, 'Are you concerned about air pollution?'. A chi-square test can be used to determine if there is a statistical significance of difference between the actual answers and what might be the expected answers. The chi-square test can also be used to determine whether or not there is a relationship between two variables, such as the development of an attitude or value and being in the programme or not in the programme. See Appendix B for directions on doing the chi-square test.

INTERPRETING THE RESULTS OF YOUR ANALYSIS

After you have performed your analyses, summarized your results descriptively, carried out any tests of statistical significance, and determined practical

significance, you are ready to draw conclusions and make inferences. You are cautioned against making broad, sweeping conclusions. Rather, err on the side of conservatism giving consideration to the following guidelines:

1. Be aware of the strengths and weaknesses in your designs. For example, to what extent can outside influences be ruled out?

2. Take into account the strengths and weaknesses of the instruments. For example, what proportion of the data is objective fact and what proportion is subjective opinion?

3. Acknowledge the strengths and weaknesses of the analysis techniques.

4. Discuss the results of the evaluation programme with your students.

5. Interpret and justify the results in terms of achievement of objectives and unanticipated outcomes.

6. Describe strengths and weaknesses of components of the instructional programme and give supporting evidence for their successes and failures.

7. Consider the appropriateness of the programme rationale, goals and objectives in light of the results.

8. Make any necessary qualifications.

9. Avoid making claims which should not be made.

CHAPTER IV: HOW CAN I USE THE RESULTS OF MY EVALUATION? (STEP FOUR)

In Chapter I, three important benefits of evaluation were identified: (1) improvement of the programme; (2) greater growth in student learning; and (3) greater support for the programme. Critical to realizing these benefits are reporting results and improving the instructional programme - the final two stages of the evaluation process. Together, they represent the major ways you can use the results of your assessment.

REPORTING RESULTS

Reporting the results of your evaluation is an essential step in your programme evaluation. Through reports you can inform others of the strengths and weaknesses of the programme and its overall quality, demonstrate a measure of accountability, generate confidence, gain support, organize information for programme improvement, and provide feedback to participants.

Identifying Audiences

There are many audiences for your evaluation report. They include:

students;

school administrative personnel;

parents;

interested citizens and groups;

other teachers;

public agencies;

funding organizations.

You may wish to place your audiences in somewhat of a priority listing in order to identify major audiences to which you will be reporting. Remember that your students represent a special audience because they were intimately involved with you in the evaluation itself. Because the data is a reflection of their response to the evaluation process as well as their performance in the programme, their rights regarding its use should be respected.

Deciding What to Say

Essentially you should address each of your programme's objectives and unanticipated outcomes, both positive and negative. Report the extent they were met or not met in clear and honest terms presenting and explaining all supporting evidence and how it was obtained. Give special attention to the rights and concerns of your students and their anonymity. You should also comment on the achievement and appropriateness of your programme's goals and the needs to which they relate.

Describe and explain those aspects of your programme that will need to be changed, dropped and kept. Final statements of summary, conclusion and recommendation usually complete the report.

Deciding How to Say It

Your report is only as useful as it is motivating and understandable. Therefore be sure that your organization, content and presentation are:

relevant to the purpose of the report;

clear and to the point;

specific and not so general that crucial points are missed.

Deciding When to Report

In planning your evaluation, you will need to decide when and how often to report. This will depend on the type of data and when it will be available, who the audiences are, and how the reports will be of use.

Interim Reports

Some audiences may need interim reports to monitor the progress of the programme. This may apply to school administrative personnel and funding agencies in particular. You, as the teacher and programme planner, will often need to prepare interim reports in order to make necessary programme changes. Formative evaluations are valuable sources of information for these reports.

Other audiences will only be concerned with final reports based upon summative data. And of course there are those who will need both interim and final reports.

Sometimes interim reports may be carried out orally and informally. If you are unsure of your audience's needs and reporting requirements, a quick check might be in order.

The Final Report

* Purpose

The purpose of the final report is to summarize the results of the evaluation, draw overall conclusions, and make recommendations.

* Outline (suggested)

(a) Abstract

1.? Write a brief paragraph outlining the major findings, conclusions and recommendations. Often busy people will only read this part of a report.

(b) Programme rationale, goals and objectives

1. Describe in easily understood terms what you hoped to accomplish and why.

2. Indicate how the programme may generate unanticipated outcomes that could become future objectives.

(c) Programme description

1. Describe who is participating in the programme, including teachers, students, other staff, support personnel, and members of the general public.

2. Describe the length of the programme and dates of operation.

3. Give funding levels and sources, if appropriate.

4. Describe major elements of the instructional programme - learning activities, special instructional resources, and learning environments.

5. Describe unique aspects of the programme.

(d) Programme evaluation

1. Explain the evaluation purposes, design, instruments and procedures.

2. Identify who was involved in the evaluation.

(e) Programme results

1. Describe positive results of successful activities in terms of student growth and impact on other people and the environment.

2. Describe those results that are unclear or marginal.

3. Describe unanticipated outcomes, both positive and negative.

4. Describe major observed changes or shifts in understanding, thinking skills, values and attitudes, and behaviour and changes observed in the environment.

(f) Conclusions

1. Summarize the extent each objective was met.

2. Summarize unanticipated outcomes and their significance.

3. Discuss the impact of the evaluation on the programme's rationale, goals and objectives.

4. Summarize the value of the programme's activities and other elements.

5. Discuss the programme in the perspective of broader needs of society and education.

6. Relate the effectiveness of the programme to costs.

(g) Recommendations

1. Recommend any changes needed in the total programme.

2. Recommend resources and other means of support for programme improvement.

3. Recommend a schedule for the implementation of changes.

* Some special considerations

1. Use graphic displays whenever the evaluation results may be shown more clearly.

2. Give special attention to the design and layout of pages and overall appearance of the report. Have an attractive cover with a short-succinct title.

3. Write clearly and to the point. Avoid unnecessary words, jargon and lengthy phrases. Organize the writing so that it flows logically and makes sense.

4. Have a draft of the report reviewed and thoroughly proof-read before the final copy is written.

IMPROVING THE INSTRUCTIONAL PROGRAMME

Ultimately the purpose of evaluation is to improve the programme so that it will yield greater returns with respect to its impact on student learning, society at large, and the quality of our environment. Perhaps the first step is to examine the recommendations of the final report in light of the programme's rationale, goals and objectives. You should be aware of any changes needed at this fundamental level and seek to implement them first. All other improvements will necessarily relate to these overall directions.

Correcting Deficiencies and Improving Growth in Student Learning

One of the more immediate uses of the evaluation results is to correct deficiencies in understanding and in thinking and manipulative skills revealed during the assessment. In this sense the evaluation may serve as a diagnostic-prescriptive tool. Many deficiencies, such as lacking an understanding of essential facts or necessary manipulative skills, might be corrected right away, far in advance of long-range programme improvements aimed at future students.

Improving the Instructional Programme

Remember that you, the teacher, have control over the instructional programme and are responsible for instituting the necessary changes called for in the recommendations. Guard against your biases. It may be necessary for you to abandon a pet teaching method, learning activity, or instructional act. But be honest with yourself and remember that the ultimate goal is to improve the efficiency and effectiveness of your programme.

Major Considerations

1. Set priorities. Decide which recommendations to implement first, second, third and so on.

2. Establish a time-line - a schedule for implementing the recommendations.

3. Identify the major opportunities and constraints related to the implementation of the recommendations. These will include the following:

time;

money;

human resources;

information;

materials and equipment.

4. Monitor the implementation process.

CHAPTER V: HOW CAN I PUT IT ALL TOGETHER? (A CASE-STUDY OF AN EVALUATION)

This chapter presents a case-study of an evaluation programme. Although portions of the example are fictional, it is based on several real-life environmental education programmes which the author has developed, implemented and evaluated. Many of the evaluation concepts and procedures presented in this guide are demonstrated in the following case-study.

A TWO-WEEK UNIT ON ENVIRONMENTAL PROBLEM SOLVING

General Description

This is a case-study of a two-week unit designed for students at the middle school level (grades six, seven and eight). The unit aimed to develop critical thinking skills in (1) environmental discovery and inquiry; (2) problem finding; and (3) problem solving. In addition, it focused on developing in students the motivation to help prevent and solve environmental problems.

During the programme, students met 80 minutes per day, five days per week for a total of ten classes. Class size was 25 students. Three classes of students received the two-week unit, which was taught three consecutive times covering a span of six weeks. The focus of the instruction was on the environment of the school site which contained about two hectares (five acres) of fields, woods, buildings, athletic fields, playgrounds, walks and drives.

A key factor in the planning, teaching and continuation of the unit was an environmental education committee. This was an interested group of individuals consisting of a teacher, the principal, the assistant superintendent, a school committee member, representative of an interested community organization, and a student.

Programme Rationale

Increasing problems of water pollution, air pollution, hazardous waste disposal, soil erosion, aesthetic degradation along with issues of land use were being experienced by the community. Citizens were increasingly being asked to vote on local ordinances, participate in public forums, elect representatives and vote in referenda dealing with the resolution of these problems and issues. Clearly there was a need to help students grasp key ecological concepts, acquire critical thinking skills, and clarify attitudes and values so that they would be able to participate effectively in efforts to prevent and solve environmental problems and resolve related issues.

Programme Goal

To produce a citizenry that is knowledgeable concerning the total environment, able to participate effectively in the maintenance and improvement of the quality of the environment, and motivated to do so.

General Objectives

(Affective objective)

1. To help individuals develop an active concern for the quality of their environment.

(Cognitive objective - understanding)

2. To help individuals acquire an understanding of the total environment, their relationship with their environment, and environmental problems and issues.

(Cognitive objective - thinking skills)

3. To help individuals to develop thinking skills necessary for the prevention and solution of environmental problems and issues.

(Action skill objective)

4. To help individuals develop necessary action skills for the prevention and solution of environmental problems and issues.

Behavioural Objectives

The following represent two major behavioural objectives in the unit.

1. (Affective objective)

The student will demonstrate a motivation to help prevent and solve environmental problems by volunteering his or her efforts.

2. (Cognitive objective - thinking skills)

The student will be able to demonstrate thinking skills related to environmental discovery and inquiry, problem identification, and problem solving.

The Learning Environment

The learning environment consisted of a classroom and the school site. Special instructional resources included a slide programme and tools and materials for investigating and mapping the site and for solving environmental problems on the site.

The Instructional Programme

Following is an outline of the instructional programme:

WEEK 1		
Monday	UNIT PRE-TEST	Formal pencil-and-paper test with slides Citizen questionnaire and school site survey (unobtrusive pre-test)
Tuesday	PHASE I: Discovery - Inquiry A. Introduction to the Unit 1. The need for environmental quality 2. The roles of citizens B. Presentation with slides 1. Natural and man-made components of the environment 2. Natural and human ecosystems concepts C. Student development of chalkboard map of school site 1. Map of shape and size of site 2. Identification of natural and man-made features - location, shape, size, amount, characteristics	
Wednesday	A. Student discussion and listing of important information to gather about site components 1. Kind 2. Location 3. Size - shape 4. Amount 5. Characteristics B. Demonstration of equipment and techniques for environmental investigation C. Student selection of topics to investigate D. Guidelines for compiling and mapping information	
Thursday	Student field investigation of natural and man-made environmental components of the site	

WEEK 1 (continued)

Friday

A. Student compilation of information, writing of reports, development of maps

PHASE II: Evaluation - Problem Identification

B. Presentation with slides

1. Criteria for evaluating the natural environment
2. Criteria for evaluating the man-made environment

WEEK 2

Monday

A. Review of environmental evaluation concepts

B. Introduction to environmental evaluation assignment sheets

C. Student field environmental evaluation of the school site

D. Class discussion - identification of opportunities for environmental improvement

---

Tuesday

A. Introduction to kinds of environmental problems

B. Student field identification and listing of problems for student involvement

PHASE III: Problem Solving

C. Introduction to the problem-solving process

Wednesday

A. Student selection of problems to help resolve

B. Student preparations for problem-solving

1. Description of the problem (why a problem, etc.)
2. Listing of alternative solutions
3. Selection of best solution - reasons, etc.
4. Planning for action
5. Listing of tools, equipment and supplies needed

Thursday

A. Student field problem-solving activities

B. Review of projects

C. Review of unit phases

Friday

UNIT POST-TEST (same as pre-test)

The Evaluation Programme

Because it was hoped that the unit would be incorporated into the regular curriculum of the middle school, the evaluation programme became an integral part of the unit. The process consisted of the following four steps or phases:

(a) Deciding What to Evaluate

At the early pre-planning stage, the environmental education committee developed the following questions to decide what information was important for school officials to decide on the value of the unit:

1. As a result of the programme, how much did students increase in critical thinking skills and motivation related to environmental inquiry, problem identification and problem-solving?

2. Were there any valuable unanticipated outcomes that resulted from the programme?

3. What aspects of the instructional unit should be kept, changed or eliminated?

These questions were the basis for decisions on what to evaluate.

(b) Planning the Evaluation

As explained previously, there were three groups of students, each group to participate in a two-week unit given consecutively over a six-week period. See Figure 17.

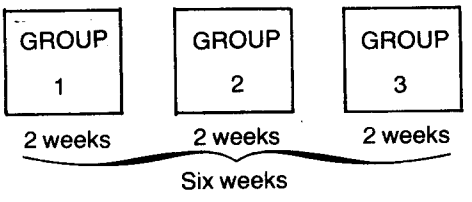


Figure 17

THREE-GROUP PROGRAMME PLAN

This arrangement made possible the following design. Group 1 became the class to pilot test the instruments and give the unit a dry run. See Figure 18.

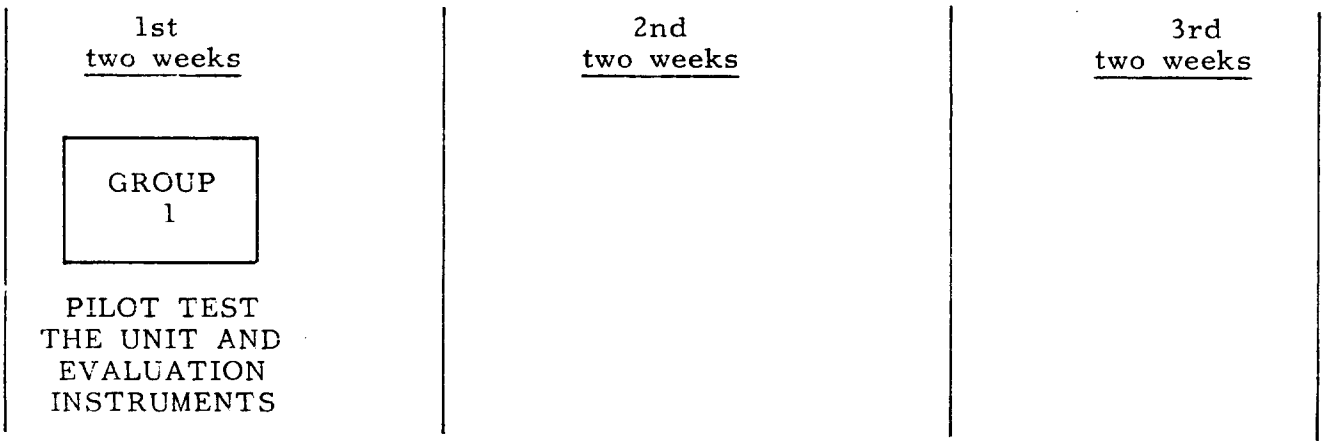


Figure 18
PROGRAMME PLAN - FIRST PHASE

Group 2 became the experimental group and received the unit during the second two weeks. During the same time, Group 3 became the control group. See Figure 19.

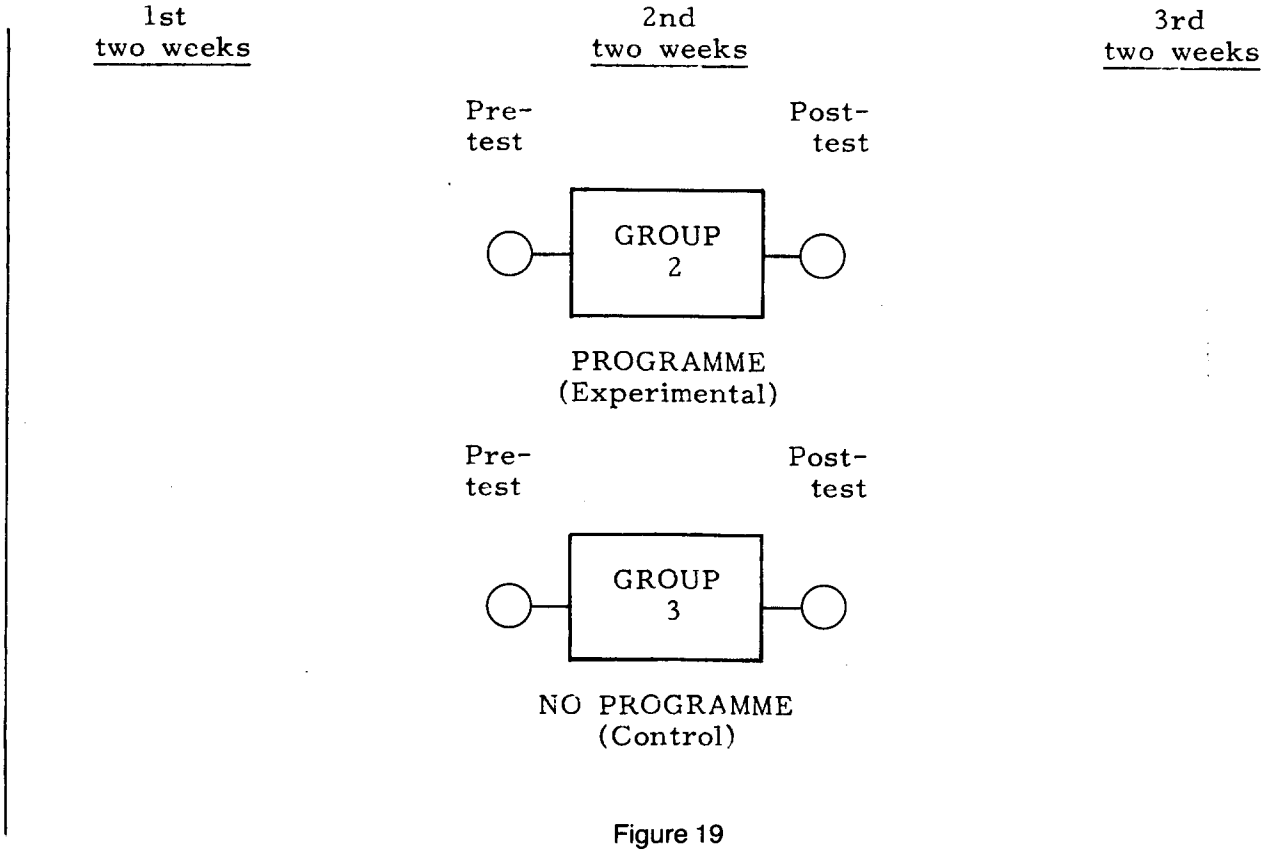


Figure 19
PROGRAMME PLAN - second phase

During the third two weeks, Group 3 received the programme. See Figure 20.

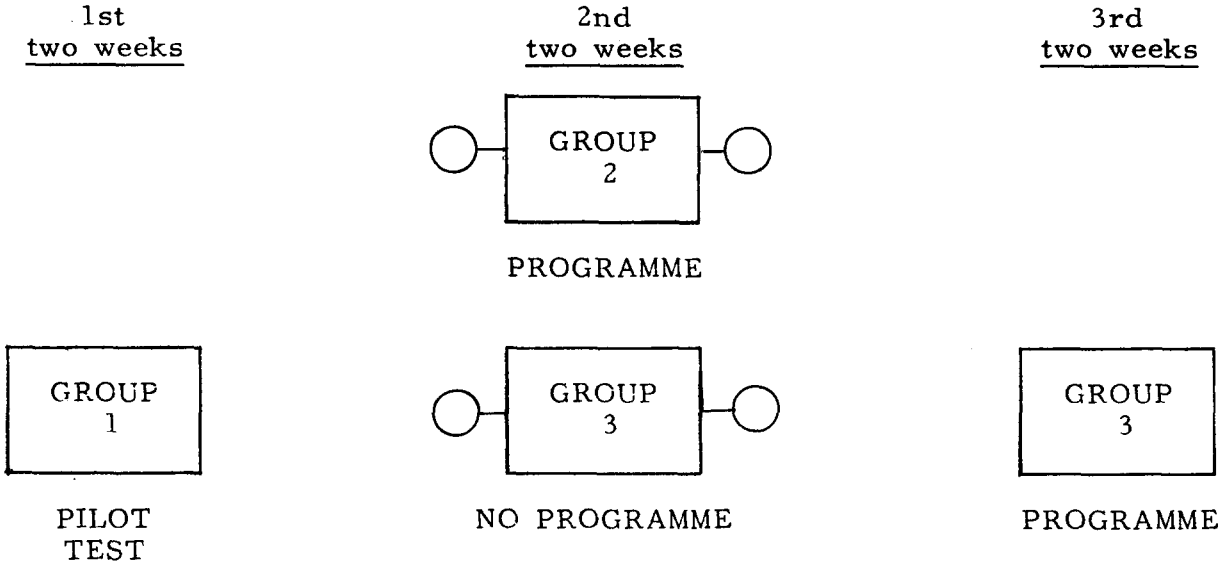


Figure 20
PROGRAMME PLAN - THIRD PHASE

Following development of the design, the evaluation instruments were planned and developed. These consisted of a pencil and paper pre-post-test with slides, two unobtrusive instruments to assess motivation, and an observation checklist to use during the unit.

The pencil-and-paper test consisted of items developed from behavioural objectives. During this part of the design process, slides were selected to accompany the items. Several hundred slides were reviewed. The slides served two purposes: (1) for some questions the slides provided a visual image about which students were asked to react; and (2) for other questions the slides served merely as stimuli for the students concerning the topic of the question.

Following the development of the instrument, it was examined by a panel of five teachers who rated each item as to the degree they believed it tested for the objective. Items that received low scores were either revised or discarded.

Following the development of the first draft of the instrument, it was pre-tested with the 25 students participating in the first two-week unit. Added to the post-test were questions about what students liked most about the unit and liked least. They were also asked to suggest changes. This second use of the draft of the test also served as a formative evaluation and provided information useful in revising the unit before it was taught the second time. Upon completion of the pre-test of the formal instrument's first draft, the test was given an item analysis and revised to increase its validity, sensitivity and time appropriateness. Also at the end of the first two weeks the first draft was given again to obtain evidence of reliability (test-retest method).

The first unobtrusive test of motivation was designed as a questionnaire from a fictitious community school site committee. The committee was supposedly in the process of gathering data on the feelings of citizens relating to the better use of school sites by both schools and community citizens. A short introduction for the

pre-test version introduced the test as a rough draft of a final version which would be developed later. Knowing that the students were participating in a school site unit, the committee was asking those students who were interested to fill out the short questionnaire as a pre-test of it before the final version was developed. The students were told that the teacher, at the request of the committee, was introducing this questionnaire at the end of the first class session of the unit. It was stressed that the teacher had no part in the committee's work and was neutral concerning it. Copies were left by the door to the classroom for students to pick up as they left at the end of the period. It was requested that they return them the next day during the second class session. During the post-test the same procedure was repeated except that this was introduced as the final revised version. The questions, however, were the same; only the introduction was different. A thank you note was attached expressing appreciation of those who helped in the development of the questionnaire the first time. All participants were asked to indicate whether or not they had participated the first time. From this test two kinds of responses could be measured: (1) the number of questionnaires completed; and (2) the nature of the responses to the questions.

The second unobtrusive test was also designed for pre- and post-test use. Students who participated were taken on a walk on a predetermined route around the school site. During the pre-test they were told that the objective was to gather information for a science unit being planned for them. For both tests each student was supplied with a three-by-five card on which there was a question about an observation he or she should make. Blank spaces were provided for responses and names. The card was folded and wrapped with an easily identified, three-coloured tissue paper band about one inch wide by ten inches long. This was attached by a piece of tape in such a way that it had to be ripped off in order that the card be opened. When the students arrived at the starting place of the survey route, the leader distributed cards to students and asked them to open them and read their assignments. He purposely ignored the students at this point to provide them with an opportunity to dispose of the wrapper in any way that they wished.

After each student was familiar with his or her assignment the group toured the site. Along the route were located 20 pieces of litter - white notebook paper, orange construction paper, white tissue paper and red ribbon. During and following the administration of the test, the following observations were made: (1) the number of wrappers discarded; and (2) the number and kinds of litter picked up.

Following the development of the unobtrusive instruments and procedures, they were also pre-tested with the 25 students participating in the first two-week unit.

The observation checklist was also developed to record the kind and extent of participation in the unit. Behaviour recorded included time on task, participation in discussions, and completion of assignments.

(c) Conducting the Evaluation

At the beginning of the second two-week session, the instructor and the student participating on the environmental education committee explained the purpose of the programme's evaluation. All pre-tests were administered during the first class session of the unit to both experimental and control groups. The post-tests were administered during the last session.

Following the administration of the instruments, data were tabulated for pre- and post-test results.

Descriptive statistical analyses were used to answer the following questions:

1. Were experimental and control groups equal at the beginning of the two-week unit?

2. After two weeks in the programme, did the experimental group score higher on the post-test than on the pre-test?

3. After two weeks in the programme, was the gain made by the experimental group on the post-test greater than the gain made by the control group's post-test?

4. Did taking a pre-test affect results on a post-test taken two weeks later?

Figure 21 is a diagram of the relationships of the questions to the evaluation design.

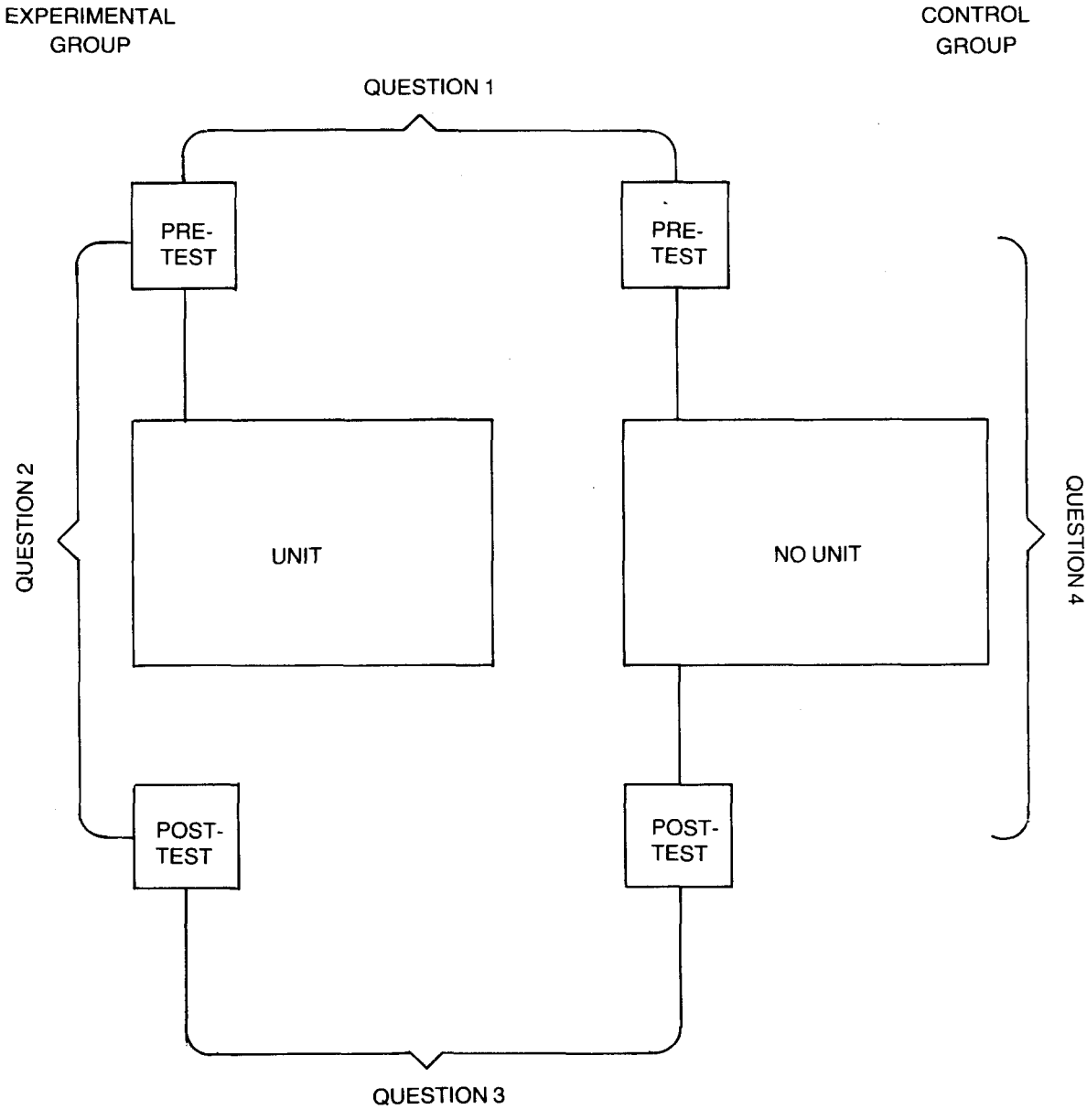


Figure 21

RELATIONSHIPS OF EVALUATION QUESTIONS TO DESIGN

The results were of particular interest. The first of these, discovery - inquiry, showed no descriptive change between pre- and post-tests of the experimental group. The second, problem identification, revealed a 7.5 mean per cent increase while the third, problem-solving, showed a 10.9 mean per cent increase. An interpretation of this might be based on the fact that the first skill is related strongly to traditional methods of education whereas the latter two represent areas to which students are not normally exposed. Therefore, one might expect greater learning to occur with respect to them. See Figure 22.

The first unobtrusive test, the questionnaire, resulted in greater participation on the post-test, thus indicating some increase in motivation to help solve environmental problems. A windstorm came up during the post-test of the second unobtrusive test and blew away the litter planted along the field trip route. Therefore, this test was invalidated indicating that unobtrusive measures are not always easy to control and thus, vulnerable to unexpected influences.

(d) Using the Results of the Evaluation

The results of the evaluation were first discussed in detail with the students. A report was then prepared, with assistance from students and the environmental education committee, and submitted to school officials. This persuaded them to continue the unit as an integral part of the curriculum.

The observation checklist and the student opinion section on the post-test yielded information for minor revisions of the unit. The improvements were incorporated in subsequent offerings of the programme.

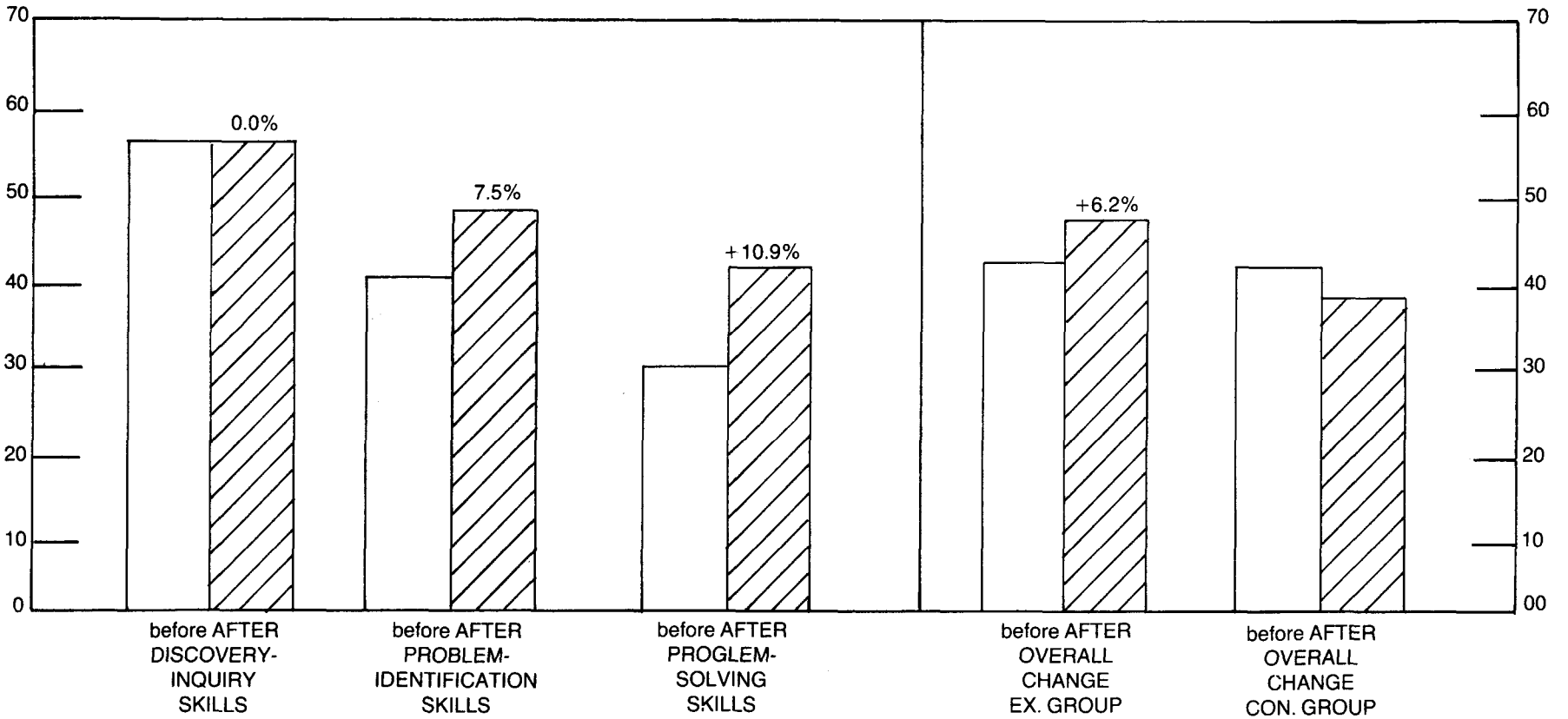


Figure 22

EFFECTS OF ENVIRONMENTAL EDUCATION LEARNING EXPERIENCES ON CRITICAL THINKING SKILLS TO HELP MAINTAIN AND IMPROVE ENVIRONMENTAL QUALITY

CHAPTER VI: WHERE CAN I GET ADDITIONAL HELP?

This chapter is a guide to further information on evaluation of environmental education. Although the availability of many good materials pertaining specifically to environmental education is limited, some excellent materials are currently being published. The sources of information included here have been selected for their availability and their potential to extend the range of your knowledge and techniques beyond the information presented in this guide.

IN-SERVICE EVALUATION WORKSHOP

The following is a suggested format for an in-service evaluation workshop:

Time	Topics
FIRST DAY	
9 a.m.-9.15 a.m.	Welcome and overview
9.15 a.m.-10 a.m.	How to avoid the end of the beginning - a case-study of evaluation
10 a.m.-10.15 a.m.	Break
10.15 a.m.-11 a.m.	What is evaluation and why bother with it?
11 a.m.-12 p.m.	What should I evaluate?
12 p.m.-1 p.m.	Lunch
1 p.m.-2 p.m.	How can I plan my evaluation?
2 p.m.-3 p.m.	How can I conduct my evaluation? (collecting and analysing data)
3 p.m.-3.15 p.m.	Break
3.15 p.m.-4 p.m.	How can I use the results?
4 p.m.-4.30 p.m.	Summary, discussion, evaluation.

Note: This basic format may be expanded, not only in content but to include opportunities to work on the development of local evaluation plans by workshop participants.

SOME SOURCES OF INFORMATION

* IOX Assessment Associates. Environmental Education, 4-9. Los Angeles, CA: IOX Assessment Associates, 1974.

This is a 148-page book containing a list of 94 behavioural objectives with related, sample test items. Answers are provided. The objectives and test items are organized into the following areas: the natural ecosystem, the human environment and environmental problems. Some K-3 objectives are included. Catalogue No. OBJ48. (For sale)

Available from: IOC Assessment Associates, P.O. Box 24095-W, Los Angeles, CA 90024-0095, United States of America.

* Iozzi, Louis A. Research in Environmental Education 1971-1980. (The first report of the National Commission on Environmental Education Research of the North American Association for Environmental Education.) 1981. 437 pp. (Order No. ED 214 762)

Available from: ERIC Document Reproduction Service, P.O. Box 190, Arlington, VA 22210, United States of America.

* ERIC Document Reproduction Service

This service has available a number of publications specifically related to the evaluation of environmental education. The following in particular are recommended for their breadth and references to other sources:

Lange, Robert R. Environmental Education Needs Assessment and Evaluation Manual. (Two volumes.) Denver, CO: Colorado Department of Education, 1980. (Order Nos. ED 199 093, ED 199 094.)

Environmental Education Program Evaluation. ERIC/SMEAC Environmental Education Fact Sheet No. 1.

* ETS Test Collection. Environmental Education Bibliography. Princeton, NJ: Educational Testing Service, 1984.

This is an annotated listing of tests currently available to test for cognitive achievement, interests, and attitudes related to the physical environment and various environmental issues. Some are for students in elementary and secondary grades, and some tests are designed to measure the interests and attitudes of all ages. A few tests are course or curriculum related. A total of 24 tests are described in the March 1984 issue. (For sale)

Available from: ETS Test Collection, Educational Testing Service, Princeton, NJ 08541, United States of America.

* The Journal of Environmental Education

Special issue - Evaluating Environmental Education Programs. Vol. 13, No. 4 (Summer, 1982). Published in the United States.

Contains a number of useful case-studies in environmental education evaluation.

FOOTNOTES

1. The author is especially indebted to Mr Ian Robottom, School of Education, Deakin University, Victoria, Australia, for many of the ideas expressed in this section.

2. Unesco. The Belgrade Charter: A Global Framework for Environmental Education (Paris, France: Unesco, 1976).

3. Anthony Kallingal. Impact of Environmental Education for Guam Schools, an Evaluation, ERIC Document ED 099 183 (Guam: Guam Department of Education, 1973), p. 31.

4. Unesco. Environmental Education in the Light of the Tbilisi Conference (Paris, France: Unesco, 1960.

5. See the following for a thorough description of these skills:
Benjamin S. Bloom, ed., Taxonomy of Educational Objectives Handbook I: Cognitive Domain (New York: David McKay Co., Inc., 1956).

6. Harold R. Hungerford and Ralph A. Litherland. Environmental Action Modules (Champaign, Ill.) Stipes Publishing Company.

7. Unesco (1980). (op. cit.)

8. Kallingal (1973). (op. cit.)

GLOSSARY

Note: The following terms are defined in reference to their use in this book.

Accountability - Answerable or capable of being explained.

Action skills - Activities involving the actual doing of investigation and problem-solving.

Affective domain - Dimension of the mind related to emotions, attitudes and values.

Anecdotal records - Accounts of noteworthy incidents and events.

Arithmetic average or mean - A measure of central tendency calculated by adding the scores in a distribution and dividing by the number of scores.

Artefacts - Refers to material objects - letters, diaries, photos, etc. - that give evidence of growth in student learning.

Assessment - Synonym for evaluation.

Attitudes - Feelings directed towards an idea, object or event. Includes likes and dislikes and perceptions of what is aesthetic and unaesthetic. Attitudes here are considered to contribute to values.

Behavioural objectives - Objectives that are more specific, performanced-based and objectively measured than general objectives. Generally consist of three parts: (1) what the student will do to indicate learning has occurred; (2) under what conditions; (3) to what extent.

Cognitive domain - Dimension of the mind related to knowing, comprehending, applying information, analysing, synthesizing and evaluating.

Content analysis - A form of analysis often used with questionnaires in which responses are categorized and frequencies determined.

Control group - The group not receiving a programme against which a programme or experimental group is compared.

Correlation coefficient - A measure of the interdependence of two variables.

Criterion referenced tests - Tests based upon behavioural objectives that establish criteria for determining levels of student growth.

Data analysis - The organization, summarization and interpetation of data using numbers, mathematical operations and word descriptions.

Descriptive statistics - Numbers that describe data.

Diagnose - To determine needs of students.

Educational variable - Refers to aspects of the instructional programme and its outcomes that evaluation programmes attempt to identify, control and measure.

Effective - The results of a programme are achieved at the desired level.

Efficient - The methods used and the setting is the best in terms of cost, time required and appropriateness for the programme's objectives.

Evaluation - Judging the worth or value of an environmental education programme - its products and its processes.

Evaluation design - A plan for what will be evaluated, who will be involved and when data will be collected.

Evaluation instruments - Tools to gather data necessary to answer evaluation questions, such as, pencil-and-paper tests, questionnaires and performance tests.

Experimental group - The group receiving a programme.

External effects - Effects of a programme on people and the environment outside of the school setting.

Final report - A report filed at the completion of a programme.

Formative evaluation - Evaluation which checks progress of a programme periodically as it is being developed and implemented.

Frequency - The number of times something occurs within a given time or in a number of possibilities.

Frequency polygon - A graphic representation of a frequency distribution.

General objectives - Objectives that describe outcomes in broad terms. They are less performance-based and more subjectively measured than behavioural objectives.

Goal - That which a programme aims to achieve. Usually establishes ideal standards and long-range direction. Broader than objectives.

Goal-free evaluation - Evaluation which looks for and analyses unanticipated outcomes.

Histogram - A bar graph.

Index of variation - A number indicating the spread or distribution of scores.

Inferential statistics - Numbers for testing hypotheses and making inferences.

Informal evaluation - Evaluation not performed with respect to a carefully prescribed set of procedures and not necessarily on a regular basis.

Instructional programme - What the teacher and students do in an appropriate setting to enhance learning.

Interim report - A report filed while a programme is still in progress.

Intuitively measured outcomes - Outcomes determined by examining one's own feelings and insights to sense something not readily evident.

Item difficulty - Refers to how hard a test item is in terms of answering it correctly.

Item discrimination - Refers to how well a test item distinguishes between high and low scoring students.

Knowledge - Knowing about something and being able to recall it from memory.

Learning environment - The settings in which learning occurs. Includes all characteristics of the classroom, the school site, the community and beyond.

Measures of central tendency - Statistical measures to detect a middle value of a set of scores.

Median - A measure of central tendency calculated as the point between the two halves of a set of scores.

Mode - A measure of central tendency calculated as the most frequently occurring score.

Monitor - To keep track of the operation of a programme and its progress.

Normal curve - A theoretical distribution of scores graphically shown as a symmetrical bell-shaped curve with most of the measurements located at the centre and very few at the extremes.

Objective - A standard of achievement which is attainable. Usually contributes towards the achievement of a goal.

Objectively measured outcomes - Outcomes determined by direct observation of evidence and uninfluenced by emotion or personal belief.

Percentage - A fraction or ratio of 100.

Performance objectives - Synonymous with behavioural objectives.

Post-test - A test given after a programme of instruction.

Practical significance - A judgement as to whether or not observed changes are large enough to be worth while.

Prescribe - To suggest ways of meeting the needs of students.

Pre-test - A test given prior to a programme of instruction.

Probability - A number expressing the likelihood of the occurrence of a specific event.

Programme rationale - Relates to the need and justification of a programme.

Proportion - A part considered in relation to the whole.

Qualitative approaches - Approaches to evaluation involving the determination of characteristics, properties, attributes. Often associated with interviews, questionnaires, observational methods, and other means of perceiving qualities not easily quantified.

Quantitative approaches - Approaches to evaluation involving measurement of number, size, degree, extent, etc. Often associated with tests and other measures easily translated into numerical data.

Range - An index of variation calculated by finding the difference between the top and bottom scores.

Reliability - Refers to the extent an evaluation instrument yields measures that are consistent each time it is administered to the same individuals.

Research - Searching, seeking out and gathering new information to further knowledge.

Response style - The tendency of respondents to choose certain responses, for example, choosing what they think everyone else will choose.

Sample - A group of individuals representative of a population of individuals.

Semantic differential - Refers to a test item used to determine where a respondent 'fits' between two opposing words describing some idea, object or event.

Simple random sample - A group in which each individual in the population has an equal chance of being included.

Standard deviation - An index of variation describing how closely the scores in a distribution cluster around the mean.

Statistical significance - The calculation of the probability that a given event could have occurred by chance. If the probability that the change occurred by chance alone is small, then the programme or some other factor might be responsible.

Student learning - Broadly defined to include the acquisition of knowledge, the clarification of values and development of moral reasoning, and the development of critical thinking and action skills.

Summative evaluation - Evaluation which determines the results of a programme at its conclusion.

Systematic evaluation - Evaluation carried out in a formal, step-by-step procedure, for example, using a detailed checklist of behaviours and pre-planned schedule to observe classroom performance.

Tabulation - To arrange a set of scores in a table or list.

Thinking skills - Ability to apply knowledge, analyse information, synthesize data, develop ideas and judge the value of ideas, objects and events.

Unanticipated outcomes - Not expected or planned in advance.

Understanding - Comprehending the meaning of something and being able to explain it.

Unobtrusive measures - Measures that do not intrude into the educational setting and students are unaware that data are being collected.

Validity - Refers to the extent an evaluation instrument measures what it is designed to measure.

Values - Feelings regarding the worth of things that have a pervading influence over lives and guide actions. Here considered to derive from attitudes.

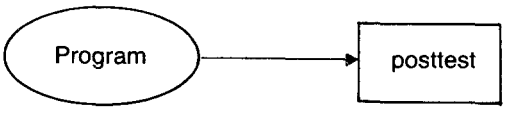
APPENDIX A: EVALUATION DESIGNS

I. DESIGNS TO ANSWER THE QUESTION: HOW MUCH GROWTH IN LEARNING OCCURRED BECAUSE OF THE PROGRAMME?

ONE-GROUP DESIGNS

One-group designs use only the students in your programme. Because the information you collect cannot be compared with other similar groups, these designs yield results that cannot be related to programme factors with as much confidence as with some other designs.

1. Post-test only design (case-study)



This design does yield some useful data on student growth in spite of weaknesses - principally the fact that you cannot compare results of the programme with levels of achievement before the programme. For example, you can ask students to indicate the extent they believe they have increased in knowledge and skill. You can also ask them to describe how their attitudes may have changed towards: (1) an environmental issue or problem; (2) their environment; (3) their role in preventing and resolving environmental problems; and (4) themselves, including their self-confidence. In this sense, this design can be used for a weak summative evaluation.

In another way the post-test-only design can yield valuable information about the programme itself. Students can be asked to reflect on what happened during the programme and offer suggestions for improvement. You, the teacher, can also summarize your own observations and opinions. These can then be compiled into a case study from which you can draw information to help in making decisions about programme improvements. In this way, the design is useful as a method of formative evaluation.

2. Pre-test/post-test design

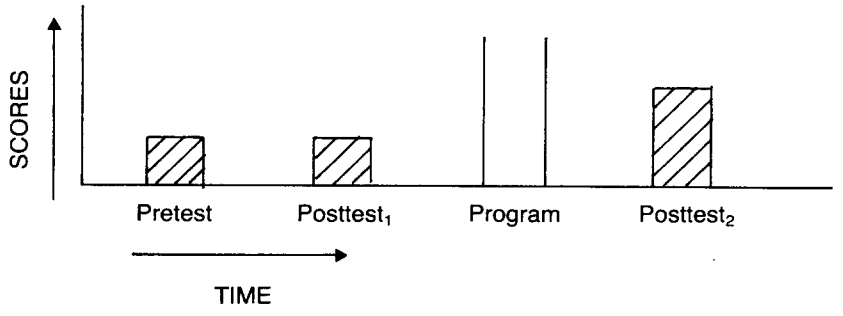
This design is an improvement over the post-test only design. However, it is still not without substantial weaknesses because it does not allow you to rule out the influence of factors, other than the programme, on the results. Some examples: outside events, such as, exposure to a book, film or television programme, could have accounted for some of the results; in the case of a long period of time between pre-post-tests, the students could have naturally matured; the use of the same tests could have influenced the results; the use of different tests may

have had some effect; or the group of teachers may have been in some way unusual. To some extent you can counteract these problems by keeping the time between tests short, monitoring outside influences, and using the same tests but withholding scores until after the post-test.

In spite of its faults, the design does offer the opportunity to compare information on learners taken after the programme with that taken before the programme. And it is particularly useful in formative evaluation - assessing the programme along the way and providing a basis for making changes. However, it cannot match the strength of preferred two-group designs to be discussed later.

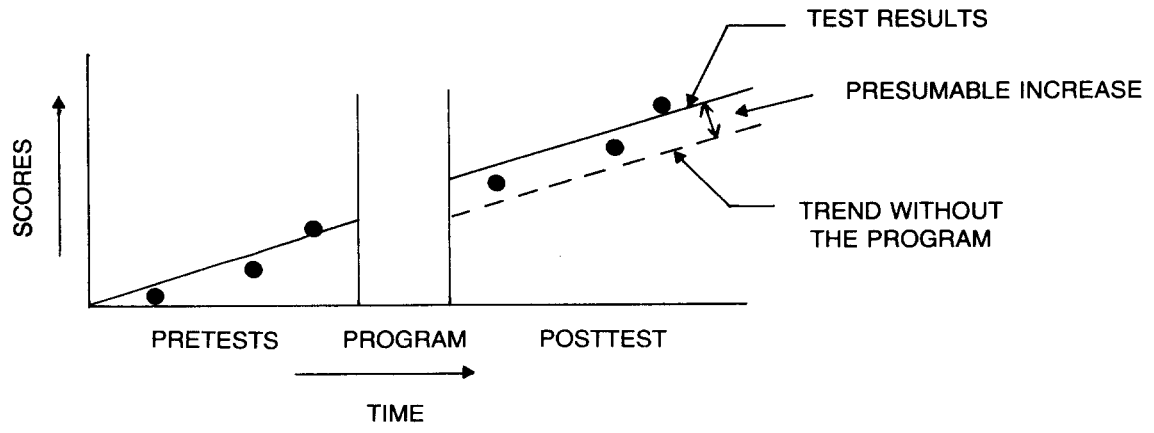
3. Flip-flop design

In this design, your class acts as its own control group. It, of course, has the same weaknesses as the previous design. In fact, if the same test is used three times the likelihood of its effect on the results is increased. And because it may lengthen the time involved, there is increased opportunity for outside events to influence the results. But it does present the opportunity to compare changes with and without the programme, even though there are, admittedly, weaknesses. Data from this design, if presented graphically, might look like that below:

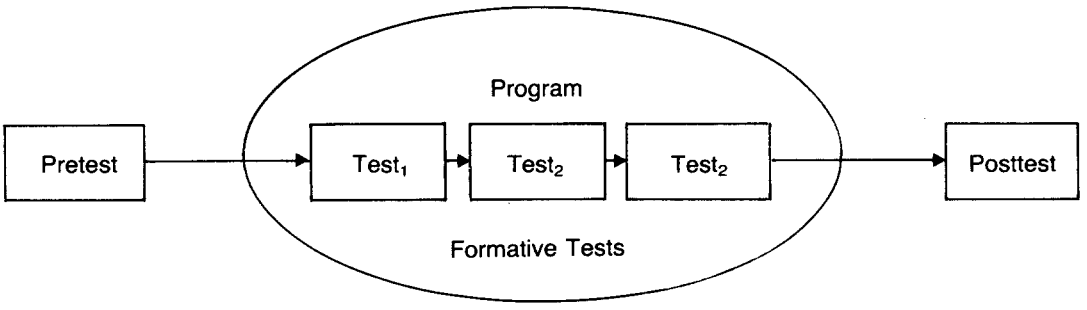


A variation of this design, sometimes called a single-group time-series design, measures student characteristics at regular intervals before and after the programme. If the programme appears to disturb the trend of results from the tests in a positive manner, this may be evidence that the programme has been effective.

A visual presentation from data produced by this design might look like the graph below:

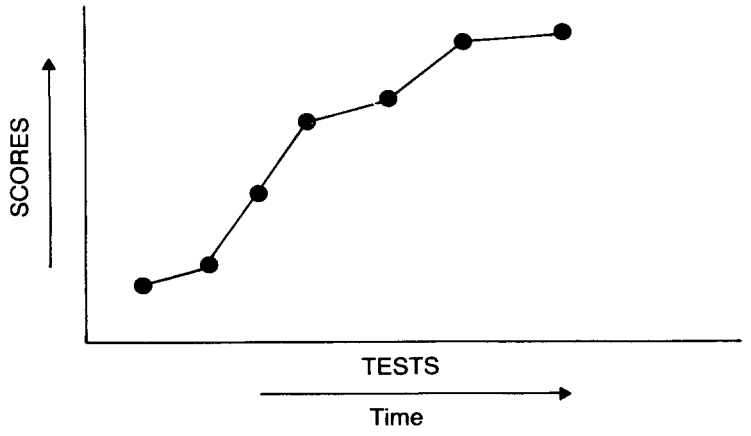


4. One-group trend analysis



This is similar to the time-series described in the previous example, except that the series of tests are given just prior to the programme, during the programme and immediately following it. This is especially useful when you have a long-running programme, say for one or more years. However, it does have the weaknesses previously described for one-group designs and you will need to explore the possibility of alternative explanations.

A graph of the trend in growth might look like that shown below:



TWO-GROUP DESIGNS

Two-group designs have the advantage of allowing you to compare growth in programme students with that of non-programme students. These designs can give you more confidence in your measurements and their interpretation. For example, they allow you to rule out the effects of outside influences. The group that receives the programme is called the programme or experimental group; the non-programme group is usually called the control group.

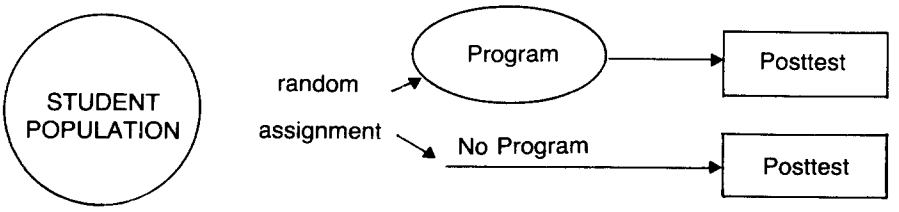
* A word on assigning students to groups randomly

When you are comparing two groups, it is important that they be as nearly alike as possible. An accepted way of increasing the probability of this is by choosing students to the programme group and non-programme group randomly.

In a simple random sample each student in the population has an equal chance of being selected. An accepted method involves using a table of random numbers such as that found in any statistics book. Assume a population of 50 students with 25 to be randomly assigned to the programme group and the remaining to be in the non-programme group. Start with an alphabetical list. (This is usually free of bias.) Number all students 01 through 50. Decide where you will start reading the numbers in the table (upper left corner, for example) and the direction you will go (across - left to right and row by row) and use the first two digits of each group of numbers until 25 students are selected. Obviously you would ignore any numbers over 50.

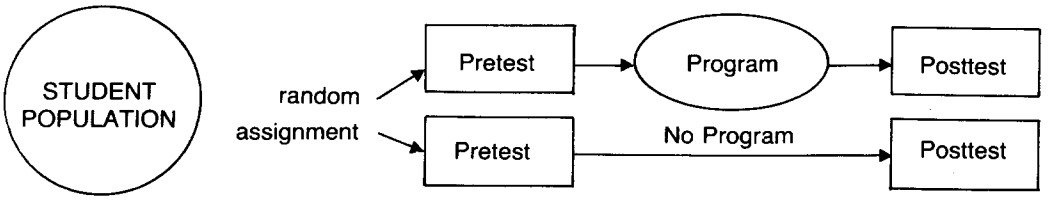
Since you are using relatively small groups, you may wish to check the distribution of any characteristics you think are important, for example, sex; if you have 23 boys in one group, you may wish to draw another example.

(1) Post-test-only control group design



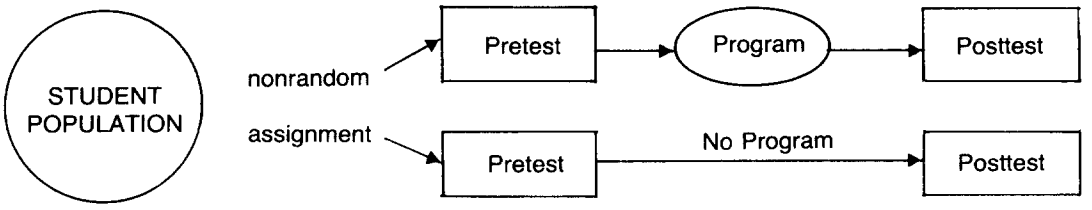
Sometimes the pre-test can be eliminated and perhaps should be, for example, if new and unfamiliar skills or concepts are being introduced, if it is felt the pre-test will influence the post-test results, or if you feel it necessary to avoid too much testing.

(2) Pre-test/post-test random control group design



This design is one of the stronger designs eliminating many threats to validity. Probably the major problem with this design is the possibility that the pre-test might influence the post-test. It is also wise to compare the pre-tests of each group to check that significant differences do not occur, for example, that the averages of pre-tests are similar. For those who are familiar with inferential statistics, both the pre-test means and the post-test means can be compared by using the t-test to demonstrate statistically significant differences.

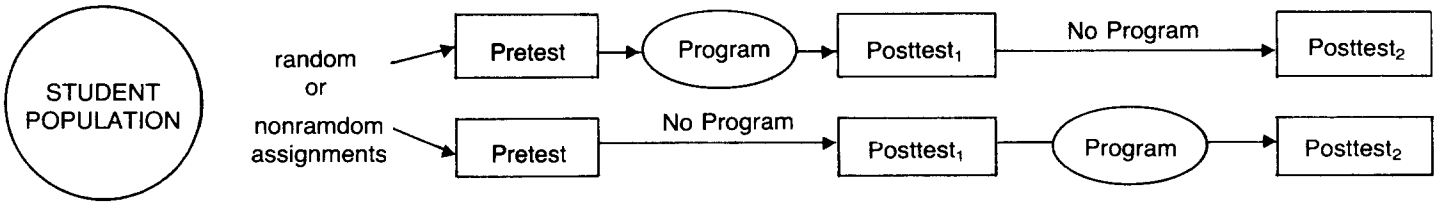
(3) Pre-test/post-test non-random control group design



More often than not, in the practical world of the school, you will not have the opportunity to have randomly assigned students in your classes and, even less, the luxury of a randomly assigned control class. In fact, you will be fortunate to be able to use another teacher's class as a control group.

If the pre-test shows that the two groups are quite different, then you can select out the scores of learners that are at variance with the group and not analyse them when you compare post-tests with pre-tests. Another possibility - you could match the groups more closely by comparing students on factors relevant to post-test performance, such as aptitude and grade-point average.

(4) Pre-test/post-test control group flip-flop design



When you are using two such groups, one group need not necessarily be left out of receiving the programme. In schools where all students are to receive the programme, say a two-week unit, half could be randomly selected as the programme group and half as the non-programme group. Both groups could be pre-tested, the two-week unit taught to the programme group, and both groups post-tested (see Chapter VI). Then the control group could receive the two-week unit and both groups receive a third test. The average score could be compared three times. The pre-test scores should be nearly equal for the two groups; the first post-test scores should show the programme group higher; and both groups should be nearly the same on the second post-test results - although there might be some forgetting.

II. DESIGNS TO ANSWER THE QUESTION: WHAT ASPECTS OF THE INSTRUCTIONAL PROGRAMME OR LEARNING ENVIRONMENT CONTRIBUTED OR TOOK AWAY FROM GROWTH IN STUDENT LEARNING?

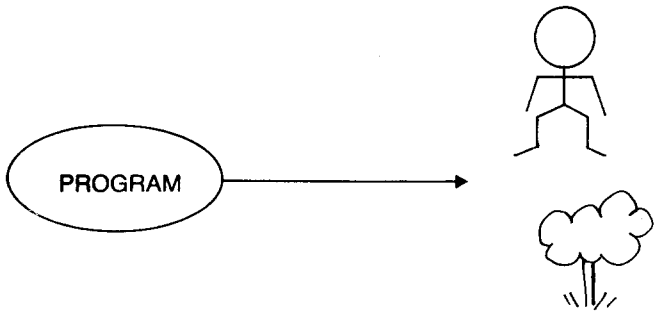
There may be times when you would like to test the effect of a new teaching technique, such as a simulation game or field trip, or a new resource, such as a filmstrip or learning centre. Ideally you should compare results from a group using the new technique or resource with the results from a group not using it.

All of the two-group designs described could be used to evaluate the effect of specific teaching techniques, some aspect of the learning environment, or a learning resource on cognitive, affective and skill development. The one-group designs may also be used, recognizing, of course, those weaknesses previously described.

When evaluating the teaching techniques or the learning environment, you are interested in more than their effects on growth in cognitive, affective and skill areas. You are also interested in the opinions of students - their interest, motivation, enthusiasm - for the innovation itself. In this case a questionnaire, interview, or other data collection method can be used as a part of the post-test in all the designs. And with all the designs you should make careful observations of student responses during the time they are involved in the activity, with the resource, or in the learning environment. Such observations would actually be a kind of formative evaluation.

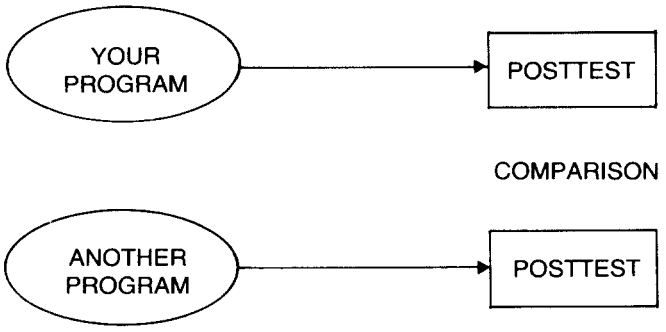
III. DESIGNS TO ANSWER THE QUESTION: HOW DID THE PROGRAMME AFFECT OTHER PEOPLE AND THE ENVIRONMENT?

To answer this question practically, your evaluation would fall into the category of a post-test design. In the case of other people, for example, those in the community who have benefited from work to improve a park, this would, most likely, take the form of a survey questionnaire or interview.



To determine direct effects on the environment you, your students, and other people could complete opinionnaires and reports. In some situations, you might also consider long-range follow-ups.

Designs to answer the question: How do the results compare with those of similar programmes?



There may be opportunities for you to gain a broader view of your students' growth by comparing the results of your programme with another programme. The comparison will mean more if you can show the similarities between the two groups of students, the programmes and the settings. Following are several important considerations:

student age, reading level, ability, male/female ratio, prior achievement;

school and teaching philosophies;

school size and grade levels;

teacher training and experience;

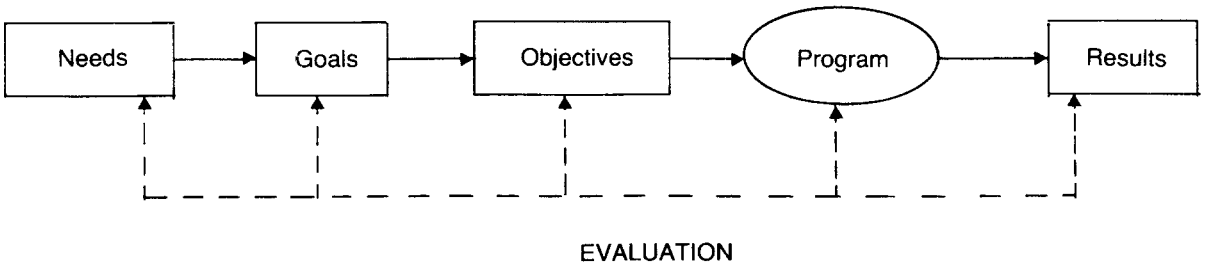
community size and population;

cost of education per student;

family income (median);

programme goals and objectives.

IV. DESIGNS TO ANSWER THE QUESTION: IS THE PROGRAMME'S RATIONALE VALID AND ARE THE GOALS AND OBJECTIVES APPROPRIATE?



For the long-range benefit of your programme and future generations of students who receive it, perhaps the real importance of evaluation lies in how you use the results to improve your programme. This means re-examining your evidence and assumptions of need (programme rationale) in light of your evaluation findings. And it means looking at the appropriateness of your goals and the practicality of your objectives. For example, if your goal is to have all students select environmental careers, you may find, after teaching the programme a few times, that this goal is unrealistic and should be modified.

Some General Evaluation Design Considerations

Time spent in programme activities

Be aware of the time your students are actually involved in the programme. Watch for absenteeism. Check for differences between programme and non-programme groups.

Comparability of control and programme groups

Take as much care as possible to ensure that the two groups are comparable on the factors you will measure and other factors, such as ability, etc., that will influence your outcomes.

Maintenance of the programme's integrity

Monitor the programme and evaluation activities to be sure that they are being implemented as intended and honestly.

Student and teacher attitude

Poor attitudes towards the programme can undermine any effort no matter how well intended its designers. Be aware of biases and continue to work to remove them. Continuously check to see how it is being received - how interested, enthusiastic and supportive its participants are.

Inappropriateness of sampling within classrooms

Sometimes when a great many students are involved in a programme, it is to your advantage to only measure the effects of the programme on a few and then generalize to the total population. It is much less time-consuming and expensive to do so. However, unless you have a large school or district, sampling is not practical. For example, it has been found that if you have as many as 100 students in the programme, you will need to sample 80 to be sure that you are within five percentage points of what the results would have been had all 100 been tested. This is with a 90 per cent certainty. In a class of 25 you would need to sample 24. However, if you have 1,000 students in the programme you would only need to sample two hundred and seventy-eight. Again refer to statistical references for specific guidance.

APPENDIX B: INSTRUCTIONS FOR THE CHI SQUARE TEST

Some measurements of your programme may yield results causing you to wonder if they are merely what might be expected or are a result of something about your programme. For example, suppose you want to find out if showing a film on environmental pollution is going to make a difference in the number of students passing an achievement test on environmental pollution. You show the film to half of your class but give the test to the whole class following the showing. You find that among the 15 students seeing the film nine passed the test and six did not and, of the other 15 who did not see the film, five passed the test and 10 did not. Now, did the film really make a difference or did your results occur by chance? The chi square test can be used to help you to answer this question.

First, construct the following table:

	Passed	Did not pass	Total
Saw the film	9* (7.0)	6 (8.0)	15
Did not see the film	5 (7.0)	10 (8.0)	15
Total	14	16	30

Second, compute the expected frequency for each cell in the table. The expected frequency is obtained by multiplying the total in the row to which the cell belongs by the total of the cell's column and then dividing by the grand total. For example, to compute the expected frequency of the cell identified with an asterisk * above, multiply 14 x 15 = 210 - 30 = 7.0. This number is placed in parenthesis in the cell. As you can see, the expected frequencies are the same for these who passed the test whether or not they saw the film. The same is also true for those who did not pass the test. This makes sense because, if the film made no difference on the test, students would be expected to do the same, whether they had seen the film or not.

Third, compute chi square (X^2) using the following formula where fo = the observed frequency and fe = the expected frequency:

$$X^2 = E \frac{(|fo - fe| - 0.5)^2}{fe}$$

$$X^2 = \frac{(9 - 7.0 | - 0.5)^2}{7.0} + \frac{(|6 - 8.0 | - 0.5)^2}{8.0}$$

$$+ \frac{(|5 - 7.5| 0.5)^2}{7.5} + \frac{(|10 - 8.0| - 0.5)^2}{8.0}$$

$$X^2 = 0.32 + 0.28$$

$$+ 0.32 + 0.28$$

$$X^2 = 1.20$$

Note: When your tables have more than four cells (2 x 3 or greater) the 0.5 may be removed from the formula. Thus the formula would be:

$$X^2 = E \frac{(fo-fe)^2}{fe}$$

Fourth, determine degrees of freedom (df). The formula is as follows: df = (r-1) (c-1) where r equals the number of rows and c equals the number of columns. Hence, for the table above: df = (2-1) (2-1) = 1x 1 = 1 df.

Fifth, refer to a table showing the significance levels for X^2 (found in almost any statistics book). A portion of such a table is reproduced below:

TABLE OF SIGNIFICANCE LEVELS FOR X^2

Degrees of freedom	Significance levels							
	.50	.25	.10	.05	.025	.01	.005	.001
1	.5	1.3	2.7	3.8	5.0	6.6	7.9	10.8
2	1.4	2.8	4.6	6.0	7.4	9.2	10.6	13.8
3	2.4	4.1	6.3	7.8	9.4	11.3	12.8	16.3

One can see that a X^2 of 1.2 for one degree of freedom gives a significance level of .25. This means that 25 times out of 100 the observed difference in the test results of students seeing the film was due to chance. Or, alternatively, there is a 25 per cent chance that the difference occurred because of a random fluctuation. In other words you would probably be wrong 25 times out of 100 if you said the film made a difference. This is a high risk; most evaluators would rather have the effects of chance reduced to, say, a .05 significance level or less.

Finally, some rules to follow in using X^2:

1. The data in the cells must be frequencies - numbers of subjects or events in categories.

2. Each subject or event in a category can only be counted once.

3. In 2 x 2 tables there must be at least five tallies in each cell; in large tables, there must be two or more tallies in each cell.

INDEX

Coordination LAB/84

DOCUMENTS IN THE ENVIRONMENTAL EDUCATION SERIES

No.			Year
1	Trends in Environmental Education since the Tbilisi Conference	Ar, E, F	1983
2	Guide on Gaming and Simulation for Environmental Education	Ar, E, F	1983
3	Module éducatif sur la conservation et la gestion des ressources naturelles	E, F	1983
4	Educational Module on Environmental Problems in Cities	E, S	1983
5	Environmental Education Module for Pre-Service Training of Teachers and Supervisors for Primary Schools	Ar, E, F, S	1983
6	Environmental Education Module for In-Service Training of Teachers and Supervisors for Primary Schools	A, E	1985
7	Environmental Education Module for Pre-Service Training of Science Teachers and Supervisors for Secondary Schools	Ar, E, F	1983
8	Environmental Education Module for In-Service Training of Science Teachers and Supervisors for Secondary Schools	E, S	1983
9	Environmental Education Module for Pre-Service Training of Social Science Teachers and Supervisors for Secondary Schools	Ar, E, F	1985
10	Environmental Education Module for In-Service Training of Social Science Teachers and Supervisors for Secondary Schools	E	1985
11	L'énergie : thème intégrateur en éducation relative à l'environnement	E, F	1985
12	*Guide on Environmental Education* Evaluation at School	E	1985
13	Guide on Environmental Education Values Teaching	E	1985
14	Interdisciplinary Approaches in Environmental Education	E, F	1985
15	A Problem-Solving Approach to Environmental Education	Ar, E, F	1985
16	Environmental Education Module on Desertification	F	1985
17	A Comparative Survey of the Incorporation of Environmental Education into School Curricula	Ar, E	1985
18	The Balance of 'Lifekind': An Introduction to the Human Environment	Ar, E	1985
19	Analysis of Results of Environmental Education Pilot Projects		1985
20	Environmental Education: Principles of Teaching and Learning	F	1985
21	Environmental Education Module on Health, Nutrition and the Environment		1985
22	Procedures for Developing an Environmental Education Curriculum	E, F	1986
23	Guidelines for Developing Non-formal Environmental Education	E, F	1986
24	Environmental Education in Technical and Vocational Education	E, F	1986
25	Strategies for the Training of Teachers in Environmental Education	E, F	1987
26	Environmental Education: A Process for Pre-service Teacher Training Curriculum *Development*	Ar, E	1988
27	An Environmental Education Approach to the Training of Elementary Teachers: A Teacher Education Programme	Ar, E	1988
28	Environmental Education in Vocational Agriculture Curriculum and Agriculture Teacher Education in Michigan, U.S.A., A Case-Study	E	1988