A SYSTEMS APPROACH TO PHYSICS EDUCATION AT FIRST YEAR UNIVERSITY LEVEL

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Doctor of Philosophy
at the
University of Waikato
by
ROGER JOHN OSBORNE

University of Waikato

1976

ABSTRACT

The research work described in this thesis evaluates the contribution which a systems approach can make to the development of an instructional programme. The thesis investigated is that such an approach can lead directly to a better understanding of the instructional process in a given subject and thus indirectly to more efficient instruction in that area.

The term systems approach is first reconsidered and a view is established which is broader than that normally considered in education. Two distinct aspects of this systems approach are identified—an "applied" aspect and a "basic" aspect. With respect to curriculum development and analysis, the *applied* aspect involves the design, development and evaluation of a learning programme. Here evaluation includes the development of specific instruments and procedures which can lead directly to the improvement of the learning programme. The *basic* aspect involves the investigation of the developed learning system in terms of models which explore the interactions between an instructional programme and the students involved.

These two aspects of the systems approach are then applied to the development and analysis of an instructional programme in physics at the first year university level. The determination of objectives, the design of instruction and the evaluation of instructional methods are detailed. Since the objectives of the instruction are generalized learner objectives, particular consideration is given to the measurement of student attainment with respect to these objectives. Four distinct aspects of student attainment are identified by the use of principal components analyses.

In terms of the basic aspect of the systems approach, both general diagrammatic models and specific statistical models are developed to obtain a better understanding of the particular first year university learning system under consideration. Various statistical models, which attempt to establish and clarify interrelationships within the instructional programme are explored. Path analysis is used as the major statistical technique.

Through the application of the systems view to the development and analysis of the first year university physics instructional programme, it has been possible to see some of the potential, some of the problems, and some of the limitations of the systems approach to physics education. In particular the systems approach, as considered in this thesis, appears to provide a broad and comprehensive framework on which to base curriculum work and to provide a viewpoint which leads to an emphasis on particularly important aspects of an instructional system—namely the objectives, the teaching methods, the assessment procedures and the feedback mechanisms within the system.

ACKNOWLEDGEMENTS

I wish to express my indebtedness to my supervisors, Professor P.S. Freyberg and Professor B.S. Liley, for their continued help and encouragement, and to my colleague, Dr. A.R. Pepper, for his valued criticism and comments. I am also grateful to Professor K.W. Keohane of the Centre for Science Education, University of London, who so willingly provided the resources, and the peaceful study leave environment, which I needed to complete this research. In this regard I would particularly like to thank those of the Chelsea College Computer Centre for their unfailing assistance. Finally, I wish to express my sincere thanks to my wife, Alison, and the boys, for their patience and tolerance.

TABLE OF CONTENTS

PART C THE DERIVATION OF SPECIFIC SYSTEMS MODELS

VOLUME ONE

CHAPTER 1

INTRODUCTION

1.1 The framework of the study

The aim of the research project to be described here, was to evaluate the contribution which a systems approach could make to the development of an instructional programme. The thesis to be investigated is that such an approach can lead directly to a better understanding of the instructional process in a given subject and thus indirectly to more efficient instruction in that area.

An instructional programme in any subject may be viewed as a system with inputs (e.g., information about students and about practical constraints) and outputs (e.g., a student's knowledge, skills and attitudes). Teachers attempt to develop learning programmes which will optimise the system in terms of the aims of their teaching. The present study is concerned with an attempt to apply the systems approach to the development and analysis of an instructional programme in physics at the first year university level.

An essential aspect of the systems view is that the system be comprehended as a whole and broad questions such as "What should the objectives of a particular programme be?", "How can these objectives be achieved?", "How can a learning programme be evaluated and reviewed?" and "Can models of the instructional system be useful?" are all of prime importance.

A systems approach to instructional programmes can be considered from two distinct aspects—an "applied" aspect and a "basic" aspect.

The "applied" aspect involves the design, development, and evaluation of a learning programme. Here evaluation includes the development of specific instruments and procedures which can lead to the improvement of the learning programme, as well as the assessment of individual student performance. Hence the applied systems approach may lead to the development of a reasonably well understood educational process or system.

The "basic" aspect involves the analysis of this developed system in terms of models. For example, models can be developed in an attempt to further the understanding of the interactions between the instructional programme and the students involved. Models which are developed may well be useful in other comparable learning settings.

The report which follows has been divided into four parts:
PART A deals primarily with the applicability of a systems approach to education. The systems approach, as interpreted in this thesis, is described and related to curriculum development. Trends in science teaching and educational research are discussed in terms of different aspects of the systems approach. Finally, general diagrammatic models of a first year university physics course are constructed to act as a basis for the development and analysis of a particular instructional programme.

PART B shows how the principles described in PART A were applied in the development of a specific programme. The determination of objectives, the design of the instruction, the measurement of student achievement, and the evaluation of instructional methods and course effectiveness are detailed.

With reference to the specific instructional programme developed using the applied systems approach, PART C considers the more "basic" aspect of the systems theory. Various statistical models, which attempt to establish and clarify interrelationships within the instructional programme, are explored. Path analysis is used as the major statistical analysis technique.

In PART D, specific examples are cited of how ideas developed in this thesis have been, or could be, applied to senior secondary school physics teaching and finally the contributions of a systems approach to physics education at the first year university level are evaluated.

For ease of reference, the tables and figures referred to in the text are included in Volume Two in the relevant order. Also included in this second volume are the bibliography, and appendices giving more detailed information than is warranted in the main text.

1.2 The setting

In 1969 a School of Science was established at the University of Waikato and undergraduate teaching in science, including physics, commenced in March, 1970. The two first year physics courses, which together form the instructional programme considered in this thesis, were developed and analyzed during the period 1970 to 1974 inclusive.

The establishment of the Physics Department provided a unique opportunity to develop completely new physics courses and the overall structure of the three year degree programme in physics was designed by Professor B.S. Liley, Head of Physics. The author of this thesis has had overall responsibility for the two first year courses since 1970. In Appendix A further background information is provided. Details of the New Zealand educational system, the University of Waikato degree structure, as well as very early decisions about the structure and general aims of the first year courses in physics are given.

In the period 1970 to 1974 all lecturing and almost all the tutoring and laboratory supervision in the first year physics courses were carried out by the foundation academic staff of the department (Professor Liley, the author, and Dr. A.R. Pepper). In 1973 Mr. K.M. Nicholas, a senior secondary school physics teacher, was the Secondary Schools' Teaching Fellow within the Department of Physics. He was not only involved in tutoring and laboratory supervision, but he also assisted with certain aspects of the research described in this thesis.

PART A

THE APPLICABILITY OF SYSTEMS THEORY TO EDUCATION

CHAPTER 2

THE SYSTEMS APPROACH

OUTLINE: The systems approach is analyzed and two distinct aspects of the approach are identified. Also a clear distinction is made between the systems approach in the physical sciences and in the social sciences. Consideration is then given to the applicability of the systems approach to education in general and curriculum development in particular. From this viewpoint various components of curriculum development and analysis are identified and placed in perspective.

2.1 The systems approach and science

General systems theory (G.S.T.), as advocated by Bertalanffy (1968) and others, has as its central theme the abstraction and generalization of mathematical models for indivisible systems of interacting particles. The major functions of G.S.T. research include;

(i) investigating the isomorphism of concepts, laws and models in various fields and assisting in the useful transfer from one field to another,

(ii) encouraging the development of adequate theoretical models in the fields that lack them, and

(iii) minimizing the duplication of theoretical effort in various fields by promoting communication amongst specialists.

It can be argued that G.S.T. is trivial because the isomorphisms are simply examples of the fact that similar mathematical equations can be used to describe quite different situations and that G.S.T. is misleading as the superficial similarities it emphasizes may camouflage actual differences and so may lead to wrong conclusions. However undoubtedly, proponents of G.S.T. have contributed significantly to the increased awareness in recent years of analogous problems in widely different sciences, and to the growth of a number of other subject areas which can be considered to be based on a systems approach. Together these latter subject areas provide a means of developing laws, models, and theories with interdisciplinary applications for investigating systems not understandable by investigation of the various parts in isolation. They include computer simulation, cybernetics, multivariate statistical analysis, theory of automata, game theory, information theory, and queuing theory.

In many fields systems exist but little is known about them and their structure cannot be determined directly. The problem is to ascertain the behaviour of the system empirically and, if possible, infer its structure from its behaviour. The systems approach identified in the previous paragraph is oriented to this type of problem, towards understanding an existing system in terms of models and theories. However, from a broad viewpoint this can be considered as just one aspect of the systems viewpoint and in this study it will be called the basic systems approach (sometimes called systems science). The other aspect of this broader systems view, which will be called the applied systems approach, is associated with the development and refinement of man made systems.[1] Where a man made system does not yet exist, a synthesis may be required to design its structure so that the realized design exhibits the prescribed behaviour (systems synthesis). If the system already exists (possibly only as a design), it may be possible to determine its behaviour in terms of its known structure (systems analysis). Whereas the basic systems approach is problem orientated

---

[1] While natural systems may be open or closed, man made systems are usually open systems—exchanges do occur across boundaries and the system has inputs and outputs. Often a man made system is designed to achieve certain goals and the system may at least theoretically, be optimised to achieve these aims.

(seeks to understand), the applied systems approach is product orientated (seeks to control). The applied systems approach is emphasized in systems engineering, systems analysis, operations research, systems development and systems design. The two different aspects of the systems approach are shown diagrammatically in Fig.2.1, and as implied by Chestnut (1967) the two aspects are complementary.[2]

Current literature (for example, IEEE Transactions on Systems, Man and Cybernetics and General Systems) shows that the systems approach can provide a common viewpoint for specialists working in different subject areas. Also in interdisciplinary teaching a systems approach can be used as a unifying theme. In Britain, for example, a systems approach has been used to provide a unifying theme for interdisciplinary degrees at the University of Aston, City University and the University of Salford, while at school level an A-level syllabus emphasizing systems (Electronic systems—A.E.B.[3]) has been undergoing trials since 1974.

Before the potential value of the systems approach to social science or education can be considered, a clear distinction needs to be made between the systems approach in the physical sciences, and in the behavioural and social sciences. The systems approach is most advanced in the physical sciences and engineering because in these fields (for example, consider electronics) there are precise general laws and well developed mathematics for expressing these laws and coping with the relationships within the system. Also there are clearly defined and

---

[2] The term systems approach is only loosely defined in the literature. In this study the most general connotation possible has been used and it equates with General Systems Science as used by Keuning (1974). The actual terms basic and applied are not commonly used. The basic systems approach is emphasized in texts such as Hyvärinen (1968) and Kalman *et al.* (1969). The texts of Chestnut (1967) and White and Tauber (1969) emphasize the applied systems approach.

[3] The Associated Examining Board, Aldershot, England.

precisely measurable inputs and outputs. A systems approach in the biological sciences is often more difficult as there is a lack of general theories, mathematical methods and quantitative measures, particularly in the behavioural fields. In the social sciences, theories and laws are tentative and uncertain, largely descriptive and verbal. Compared with the physical and biological sciences, which are essentially objective and impersonal, the behavioural fields deal with elements whose behaviour is subjective, and often random and conflicting. For these reasons, quantitative deterministic models closely approximating reality are more likely to be found in the physical sciences than in the social sciences where models tend to be statistical, graphical and verbal. This variation in emphasis is shown vertically in Figure 2.1.

In the social sciences there are many instances of the basic systems approach. Examples are: the simulation of social science systems (see Shubick, 1967; Dutton and Starbuck, 1971); the widespread emphasis on multivariate analysis (see Cattell, 1966a; Firestone, 1972); the application of the transfer function from engineering to aspects of human behaviour (e.g., Truxal, 1972; Tamara, 1974). More specific examples are; the use of cybernetic models in economics (e.g., Runyan, 1971); and the application of thermodynamic theory to history (Iberall, 1974). Bertalanffy (1968) has pointed out other examples; armament races and warlike conflicts can be elaborated in terms of differential equations similar to those used in ecology, the spread of rumours can be described by a generalized diffusion equation; the flow of automobiles can be analyzed by considerations formally corresponding to kinetics and thermodynamics, and concepts of general systems, feedback, information and communication theory are being increasingly introduced into sociology. On the other hand, the applied systems approach has been most useful in designing, developing and modifying social science systems. There is a widespread application of systems design and development to aspects of social science (for example, Tuckman and Edwards, 1971; Gresford, 1972; Amara, 1972).

Some of the problems and limitations of the systems approach as it applies to social science have already been specified. There is a danger of superficial similarities camouflaging actual differences, with the result that wrong, or even morally objectionable, conclusions about a situation may result. There are difficulties arising from the undoubted complexities of social science systems. There is the problem of defining and measuring the entities under consideration. For example, the inputs and outputs are likely to be abstract social qualities for which only indicators rather than measures (Lord, 1968) are available. As Churchman (1968) has pointed out, there is the difficulty of viewing social organizational systems as dispassionately as physical systems. The views of the "humanists", who are concerned with human feelings, who maintain that human organizational systems are people and who emphasize that human values such as freedom, dignity and privacy must be considered vitally important, cannot be ignored. Finally, any human organizational system is a subsystem of a larger system and the idea of treating a subsystem as isolated (apart from certain inputs and outputs) certainly needs to be viewed with caution.

In terms of the interplay between the basic and applied aspects of the systems approach in the social sciences, there has been some concern (see Berkman, 1972) that in developing and refining systems, too little attention is often given to the basic systems approach to finding out how the system under development or refinement actually functions. Berkman considers that many systems analysts think they know what the problems are before they have a knowledge of the system at work; he maintains that it is vital to place greater emphasis on identifying system elements, on measuring outputs as indication of system performance, and on developing a clear understanding of the properties of the system.

2.2 The systems approach and education

What has the systems approach to offer in the field of education? An individual's concept of an educational system is influenced by the level of decision making, evaluation, or research in which he is involved (for example, with instructional or administrative problems, with individual learners or with national systems). The type of systems approach which can be successfully applied to the development and analysis of an educational system will depend greatly on the type of system under consideration. However, it would be expected that a systems approach to educational systems, at whatever level, should;

(i) keep overall systems goals clearly in view,
(ii) clarify and co-ordinate different aspects of development and analysis,
(iii) consider problems in relation to research in other areas,
(iv) indicate priorities for research.

In large educational systems when educational administrators are concerned with inputs and outputs which can be reduced to numbers of students or teachers, dollars or buildings, a reasonably "hard" (mathematical) systems approach may be possible.[4] With emphasis on a basic systems approach, models of the educational system have been developed (e.g., Thornstad, 1967; Weizsacker, 1967; Armitage *et al.*, 1969; Windecknecht and D'Angelo, 1975). Also the applied aspects of the systems approach—operations research, cost benefit analysis, systems analysis—have been used to study educational administration problems (see Hartley, 1968; Immegart, 1973).

---

[4] As suggested by Rappaport (1970) a "hard" definition is one that permits an unambiguous recognition of the theory defined, while a "soft" definition provides only an intuitive understanding.

In the area of curriculum development and of the evaluation of instructional and learning procedures (which is the particular area of interest in this thesis) much has been written about the use of the systems approach (see Dyer, 1969; Crys and Lowenthal, 1970; Merrill, 1971; Lerner, 1973; Rath, 1973). However, almost exclusively, emphasis has been on the "applied" aspects. A preoccupation with developing and designing systems from a systems viewpoint has led to many arguments in favour of the specification of objectives of educational systems and the designing of "optimal" systems to meet these objectives. The detailed specification of behavioural objectives, the designing of educational activities in terms of measurable learner behaviours, and the precise formulation of procedures and feedback devices all have merit. However, the previously stated limitations of the systems approach in the social sciences are also highly relevant to systems approaches in education. Often objectives and instructional methods in a learning system have been defined to extremes of specificity inappropriate to a non-physical science-non-deterministic system, as Yee (1972) has outlined.

Apart from the social sciences in general, there has recently been some realization in education (see Apple, 1972; Stowe, 1973) that curriculum developers seem to be employing applied systems science to manage their problems without first understanding the complexities of the relationships involved in the educational environment. In an earlier support of this view Atkin (1968) differentiated between the "engineering" model and the "naturalistic" model for educational research and argued that the over-simplistic view of the "engineers" had shortcomings, in particular because of the insufficient attention given to competing objectives and values. He argued for an indepth study at the classroom level; a "direct onslaught on the total educational picture"; a plea for research in naturalistic settings; a plea for putting aside "engineering" for a time and a plea for phenomenological approaches.

Glass (1972) has also implicitly emphasized the "basic" and "applied" aspects of the systems approach in his catagorization of two types of inquiry in education; elucidatory and evaluative. He states that elucidatory inquiry is the process of obtaining generalizable knowledge by contriving and testing claims about relationships among variables. When the results of elucidatory inquiry (statistical relationships, models and ultimately theories) are combined with the knowledge of a particular situation, one should obtain explanations. On the other hand, evaluative inquiry is the determination of the worth of a programme product or procedure. Glass (1972, p. 12) (in an opposite view to Atkin (1968)) argues that elucidatory inquiry in education "simply does not seem to have turned up any important, reliable, replicable relationships worthy of continued study". He suggests that this is basically because of the enormous complexity of the educational enterprise and hence our inability to uncover models or theories of wide generality. Therefore he considers we should focus our attention on evaluative inquiry and curriculum development.

In contrast to natural systems, educational systems are man-made. However, unlike man made systems in, say, engineering, the development of an educational system using an "applied" systems approach (of carefully specifying intended behaviours and "designing a structure" to achieve these prescribed behaviours) will not result in a system with entirely predictable behaviour. Rather a system is produced which exists, but which is still little understood because its structure cannot be determined directly. A "basic" systems approach to the understanding of such an existing structure may well be worthwhile. It may be true that basic theories of wide generality are unlikely to emerge from the study of a specific system; however, where the aims of the system have been clearly identified, and the procedures attempting to meet these aims are reasonably well understood, it is possible that an elucidatory inquiry (basic systems approach) may uncover some "subsurface" explanations which will aid in interpreting evaluative findings for that time and place, and in turn lead to more informed decision making. In terms of wider generality such studies are likely to uncover the problems in, and suggest procedures for, the development of better models of educational systems. Further, they may indicate likely relationships to be found in other specific systems, and suggest fruitful models by which other similar systems may be understood.

Tyler (1968) has suggested educational research should provide a basis for;

(i) planning and developing educational programmes, and

(ii) understanding the educational process or parts of it.

These two aspects of educational research relate closely to the two aspects of the systems approach outlined. Both aspects are important and should influence each other. Figure 2.2 summarizes the application of the system viewpoint of Figure 2.1 to curriculum development and analysis. It is proposed that rather than concentrate solely on a rigorous applied systems approach,[5] more emphasis should be given to isolating, measuring and understanding the relationships between the broader goals and the major procedures of a learning programme. This emphasis implicitly assumes a particular view of the instructional system (which will be called a generalized curriculum model) in which the broad goals are of major importance. Once goals, or objectives, have been determined, the applied systems approach can be used to design, develop and modify the instructional programme. This will require specific development procedures and feedback. Once the

---

[5] often with an over-emphasis on specifics, on absolute levels of behavioural outcomes, on feedback channels and detailed instructional procedures.

instruction has been developed and refined, a specific statistical or mathematical model of the system can be developed which will, hopefully, provide feedback for further specific development and which may also provide useful information for proposing future generalized curriculum models.

The above order for curriculum development is an idealization which assumes that no similar learning programme previously exists. However, if a systems approach is to be applied to further development or modification of an existing programme, models of the existing programme may be desirable. As Stone (1967) suggests, in reality the order of events cannot be ideal in that one has to break into the cycle at some point while one's knowledge of other aspects of the system are incomplete—aims must be specified without being sure that they are attainable, or even desirable—the system must be regulated without sufficient information about its characteristics to design efficient control devices—data must be collected before the important variables in the system have been identified. However, the systems approach emphasizes that all these aspects of the problem are important and should be given attention.

2.3 The systems approach and curriculum development

This thesis is concerned with the application of the systems approach to the development and understanding of a specific instructional programme—a first year physics course at a new university. As emphasized in the last section, both the developmental and model-building aspects of educational research are considered important in terms of any learning system, and these two aspects of the systems approach should be integrated as far as possible. Although Stufflebeam *et al.* (1971) were primarily concerned with educational evaluation, and hence with the applied aspects of the systems approach, their work can be extended to provide a useful framework to consider both aspects of the systems approach to learning programmes.

Stufflebeam *et al.* (1971) isolated four aspects of educational evaluation.

Context evaluation provides a rationale for the determination of goals, i.e., what should the objectives of a programme be? It identifies unmet needs and unused opportunities. In existing systems, it determines the reasons why goals are not being met. The determination of the amount of change to be effected in an existing system is also an aspect of context evaluation, i.e., is a single large change or several small "incremental" changes to be made?

Input evaluation provides information for identifying the relevant resources available to meet programme goals, i.e., what is the best way to achieve these objectives? Information is required about the potential and practicability of various learning methods to achieve objectives. Input evaluation therefore also includes evaluation of the constraints on the system and of the prior knowledge and abilities of entrant students. Input evaluation tends to be more *ad hoc* than context evaluation and the methods used depend very largely on the magnitude of the changes proposed.

Process evaluation is primarily concerned with the provision of information from the operating course which can be used to modify instruction or make programme decisions. This is particularly important in the early stage of course development, but it also applies to the ongoing "monitoring" of an instructional programme. Basically process evaluation provides feedback to input evaluation.

Product evaluation measures and interprets achievement of course objectives. Whereas input evaluation provides the specification for process evaluation, context evaluation provides the specification for product evaluation. Evaluation in each of these areas can range from "unstructured and subjective" to "structured and objective", as Frazer

(1975) has recently suggested in a slightly different context.

The above four aspects of evaluation are known as the CIPP model (Stufflebeam *et al.*, 1971) and in Figure 2.3 this model for system evaluation has been integrated into the systems approach view of curriculum development outlined in Figure 2.2.

The generalized curriculum model and the specific system model are primarily concerned with the basic systems approach, with providing an overall systems view to curriculum development and with developing models of specific educational systems. Educational models may be mental, verbal, diagrammatic or mathematical. Mental (intuitive) models of complex systems are ill defined and not communicable; assumptions are not always clearly identified; and the models cannot be manipulated effectively, as Forrester (1968) has pointed out.

Verbal or diagrammatic models are better defined and more communicable and are useful in a systems approach to education. Such models can act as a guide to curriculum development and analysis and can provide a basis on which to develop mathematical models. The advantages of mathematical models are that they tend to be unambiguous; they can be verified by using observable data; and they provide the possibility for strict deduction. However, in education there are difficulties not only in the complexity of the phenomena but also in the definition and measurement of the variables under consideration.

While a generalized curriculum model is likely to be verbal or diagrammatic and basically provides a general point of view from which to consider a specific educational system, hopefully a model of a specific system can be more mathematical. Ideally the development of a mathematical model should be a cyclic process between theoretical model and experimental observations. However, particularly in the social sciences where there may be few well developed theories, models may be developed from two distinctly different points of view (Stone, 1967). The empirically based model starts with the observations and asks how they may be related. It is usually assumed that the first attempts at relating these observations will be rather approximate, and that this will suggest improvements which may be as much in the data and the way it is organized as in the "model" itself. This type of model development is exemplified in the path analysis model developed by Jonnada and Fegley (1974). The theoretical model results from considering how in principle a problem might be solved, even if initially no data are available. This kind of theory is usually concerned with the character of the solution under fairly general conditions. Usually the problem is posed very abstractly and this offers some hope of solution if sufficiently advanced methods are used or gross simplifying assumptions made. An example of this type of model development in education is the simulation model developed by Windecknecht and D'Angelo (1975).

In developing or selecting a model there obviously needs to be some clarity as to the use to which the model is to be put—whether it is to explain or to predict. In the case of prediction the primary concern is to develop a model which will best predict a future event without necessarily uncovering underlying explanations or system structure. For example, a predictive model of success in first year physics would be concerned with determining the best combination of the most pertinent predictor variables obtainable, irrespective of what these variables might be, or what causal connection (if any) exists between them and the criteria to be predicted. On the other hand, a different model using different techniques and different variables may be more useful for the purpose of clarifying what is meant by success and of understanding clearly the factors on which aspects of this success are based. Variables might be included not because of their predictive power but rather for the insight they would be likely to contribute to the understanding of the system. Specific systems models in education are likely to be empirically based and will hopefully elucidate systems structure. In practice, apart from their ultimate aim of explanation or successful prediction, the function of any model is to polarize thinking and pose sharp questions (Kac, 1969) and, as a major aim of educational research is to clarify issues and provide information for rational decision making (Nuthall, 1974), hopefully systems models will help do this.

In Figure 2.4, the different components of evaluation and elucidation outlined in Figure 2.3 have been respecified in terms of the corresponding activities which they serve in the development and analysis of an instructional programme. Both the CIPP model and many of the statistical models which have been developed in educational research have been applied by others to extensive curriculum development programmes and to large scale (national) educational investigations respectively. The work in this thesis is concerned with combining the two aspects of the systems approach into a more unified and co-ordinated system and also with investigating what relevance this framework has for developing and improving small scale instructional programmes. In subsequent chapters the ideas just discussed will be applied to the development and analysis of such an instructional system--the first year physics courses at the University of Waikato--and, in addition, the relevance and implications of the systems approach in general will be considered with particular reference to physics education in New Zealand. How Chapters 4 to 13 relate to the different components of the systems approach to curriculum development is shown in Figure 2.4.

CHAPTER 3

SCIENCE EDUCATION

OUTLINE: Some of the trends in science teaching and in science education research are briefly discussed in terms of the various components of curriculum development identified in Chapter 2. Specific studies are cited as examples of trends and of methods of approach to curricular analysis.

3.1 Science education studies and the systems approach

In Chapter 2 a framework was developed for the consideration of educational research on specific learning systems. Before proceeding to consider the application of this systems approach to one specific educational programme in physics at the first year university level, consideration will be given in this chapter to recent studies and trends in science and mathematics education and how they can be related to the framework of Chapter 2. The studies discussed are included because of the similarity of subject area, or student characteristics, to the system being investigated in this research, or because the method of investigation is of interest.

As mentioned in Chapter 2, in any particular study the different aspects of development and research are not always clearly identifiable, nor does development take place in the idealized order, as suggested in Figure 2.4. For example, a model building study on a current educational programme may well answer, and might even have been designed to answer, specific questions related to context evaluation on a new programme about to be developed. However, such a basic systems approach is likely to be even more meaningful and fruitful when the investigator, or model builder, is familiar with the applied, or developmental, aspects of this programme, i.e., he understands the goals for which the programme was designed, and considers the relationships between these goals.

One reported example of the careful development of a system, followed by studies more orientated towards a basic systems approach, is the work done on the development and analysis of the Nuffield A-level biology programme in Britain. Much of the early work in this project was naturally concerned with development (the applied approach), with finding answers to specific questions and with developing a better product. However, subsequent investigations (see Kelly, 1972a, 1972b) were more orientated towards an "analysis" of the characteristics of the system and this research is continuing.[1]

3.2 The applied systems approach: context evaluation

There are two distinct aspects of context evaluation, first the search for opportunities and pressures outside the system to establish and modify the goals of a learning programme, and secondly the evaluation of pressures inside the system for changes in established goals. Establishing the goals of an educational programme is of major importance. Unfortunately at present minimal knowledge to support the choosing of goals, and a changing attitude towards science by students, results in a rather *ad hoc* choice of goals by many curriculum developers. Many educational programmes, or projects, at the national secondary school level (for example, in Britain) have been "Big-bang" projects (Richards, 1974); objectives or aims are decided upon, the project is developed by one or two cycles of the CIPP cycle of Figure 2.3 and then distributed to teachers for introduction in their classrooms. As implied by Deeson (1974) different course developers have different goals and this can simply lead to confusion.

---

[1] Griffin and Kelly, Centre for Science Education, Chelsea College, London.

> "Course (development) Team A says throw away textbooks; B says practical work is the only true path to scientific heaven; C returns us to formality because this, that and the other; D pushes projects; E pushes audio-visual; F pushes full free flexibility.... And so it goes on.... Let us call a halt for a few years to catch our breath; let us evaluate; let us pick and choose; praise and reject; let us not allow a rupture of the project freeze. Except for, my own project, of course." (Deeson, 1974, p. 12.)

This superficially lighthearted, but very relevant comment emphasizes the need to find ways of clarifying aims and evaluating innovations.

Some teachers view each new project as some sort of seasonal epidemic disease which they hope to survive without contracting. Some teachers without the time or resources to develop their own programme from the basis of their own objectives simply accept (or have to accept because of national policy) the objectives of the project developers and their product. Some teachers check whether or not a new project can be adapted to develop the aims they see as important before deciding to adopt it, or part of it.

However, while national projects may possibly be objective oriented and develop from specifying objectives to producing an instructional product, the classroom teacher is, and is likely to remain, largely activity oriented (Eisner, 1969b). He has not the resources to develop a learning programme purely from idealized objectives. Even at best, he must normally rely on available textbooks, equipment, tests, audio-visual material, and if necessary, modify them to best suit his own objectives. As projects come and go he will ideally see his objectives more clearly, his resources widened, and his course subsequently modified. Hopefully, even if almost intuitively, he will continually proceed around the evaluative cycle of Figure 2.3, every so often making some larger changes by choice or external requirements, but more often introducing small (incremental) changes to improve the learning programme for which he is responsible.

As Richards (1974) has pointed out, there have recently been some attempts in Britain to develop "continuous creation" type projects, rather than "Big-bang" projects. In these projects, both at school and university level, information about individual teacher's expertise is gathered and published in a form in which it can be used by other teachers. While this requires a common understanding of goals and criteria for assessing "worthwhile" ideas the emphasis on individual teachers using an applied systems approach in their own learning programmes seems desirable. When the probable lifetime of a current syllabus, or "Big-bang" project is compared with the average working lifetime of a teacher, it is clear that more emphasis than at present should be given to assisting teachers to be more aware of different objectives in their teaching, to establish their own aims, to review critically available educational programmes and to adapt and modify their teaching with purposeful awareness. At the university level, where every teacher basically develops his own teaching programme, awareness of objectives has been encouraged to some extent by the introduction of centres of research and development of university teaching and the associated teacher training of university teachers which these centres have been able to provide.

The study of Welch and Walberg (1967) preceding the development of the Harvard Project Physics in the U.S.A. is one of the few reports dealing specifically with the determination of objectives for a proposed learning programme. Welch and Walberg (1967) found teachers were strongly in agreement with the proposition that physics is needed by all students, not just the academic elite. Thus one of the important aims of the course was to meet the needs of a wide range of students. In Europe, the Council of Europe (1972) surveyed and compared the general aims of teaching in fifteen European countries. Lists of aims and the emphases placed on these aims in different countries provide a basis for choosing aims for new learning programmes. In New Zealand, Duncan *et al.* (1971) sought the opinion of informed members of society on the importance they placed on different objectives of chemistry prior to the development of a new chemistry syllabus for the sixth form.

3.3 The applied systems approach: input, process and product evaluation

Once objectives have been established the problem is to determine which learning method is likely to best achieve a particular objective—input evaluation. This is a major problem in curriculum development. Recent trends towards multi-media approaches (which may introduce an element of student choice) require even more consideration to be given to the particular teaching methods which can be provided within the constraints on the system and to the potential a particular method has for achieving instructional objectives. As yet little help can be gained from educational research on different teaching methods because results of such research are (a) often conflicting or not significant, and (b) may not be transferable to the system being developed. Often the choice of teaching method has to be based on an intuitive feeling for its potential to achieve a variety of course objectives. Teacher enthusiasm for a particular teaching method is also an important factor which must be considered. Reported investigations prior to the development of a teaching programme which are examples of input evaluation in physics education are those of Walberg (1969b) and Bridgham and Welch (1969). These led respectively to certain teaching methods and teaching practices (see Welch, 1973, p. 372) being introduced into the Project Physics course.

Recent trends in Britain in evaluation of existing courses has been towards process rather than product evaluation (e.g., see Parlett and Hamilton, 1972) and undoubtedly process evaluation can provide a variety of useful feedback. A common procedure in process evaluation is the comparison of one part of a course with another. This provides useful information for course development. For example, Johnstone

(1974) compared the relative difficulty of topics in S.C.E. O-grade chemistry and then investigated which of two alternative teaching methods was more suitable for the teaching of some of these more "difficult" topics.

Product evaluation measures and interprets attainments with respect to the goals of the course. Hence product evaluation is closely related to context evaluation. Welch and Walberg (1972) measured the attainment of the goals of Project Physics by the comparison of this course with a basically PSSC method of teaching. The goals of the Project Physics authors were;

(i) to reduce the difficulty stigma attached to physics,
(ii) to reduce the mathematical orientation, and
(iii) to show physics as an intellectual endeavour rather than as applied technology.

In a sense the goals inherently contained a comparative element, and the comparative study of Project Physics with the more traditional courses did suggest that the course was successful in achieving these goals.

Where the goals of a course are aimed at student behaviour, product evaluation is associated with student assessment. Both Kelly (1971b) and Mathews (1974) have emphasized aspects of student assessment in their evaluation of Nuffield A-level biology and chemistry respectively. At university level the increasing diversity of possible student assessment techniques is increasing the feedback and teaching potential of assessment, and affecting how the product of a course with behavioural objectives can be measured. It is more difficult to obtain "useful" feedback information from product evaluation than from process evaluation and at present evaluators at the university level tend to emphasize the latter (e.g., the Higher Education Learning Project (Physics) in Britain). While occasional reports on innovations in teaching in universities provide comparative product evaluation data (e.g., Barratt and Blake, 1974), it is more common to find innovative teaching where no formal evaluation of the product has been made (e.g., Fentem, 1974, p. 8; Mansell, 1974, p. 8). This simply reflects the problems involved. However, the general interest in student assessment must lead to useful information for product evaluation in the long term. For instance, the development of profiles of student achievement in terms of the aims of the course (e.g., Klug, 1972) is related to the problems of product evaluation.

3.4 The basic systems approach

In the physical sciences the ability to identify, to clearly define and to control variables; the ability to reproduce controlled conditions; the ability to obtain precise measurements of dependent and independent variables; and the fact that the system under consideration can be reduced to a limited number of variables; all lead to basically simple but effective deterministic models. However in education, on the other hand, the multivariate nature of the systems under study; the difficulty of measuring theoretical constructs and of reproducing controlled conditions; the problems of errors of measurement and the linearity of scales, all necessitate probablistic models. For these reasons a multivariate analysis with emphasis on the natural setting and on the non-manipulation of variables is at present the most promising means by which a model building approach to organizational systems might be made. As Cattell (1966b) suggests, one teases out the causal connections amongst the data, using "intricate non-interfering finesse". Basically this approach starts with the observations and asks how they may be related. As has been stated earlier the alternative approach is to develop a model from a theoretical or abstract basis. Currently, at least in the field of instruction and learning, it is extremely difficult to develop theoretically based mathematical models of this kind which will result in testable predictions.

Most model building in natural settings in educational research has been of a multivariate statistical inference type. Typical of large scale studies, Coleman (1966) developed indices from factorial studies and then employed multiple regression techniques. In a much smaller study Keeves (1972) in Australia used a non-interfering approach to study the interrelationships among the educational environment of the home, the classroom and the peer group, and the influence on these factors on educational achievement at the primary/secondary interface. Multivariate techniques of principal components analysis, multiple regression, path analysis and canonical analysis were used. Keeves' study can definitely be classified as a basic systems approach; he concludes;

> "my enquiry was undertaken not in the hope of finding practical answers to problems, but rather in the hope that the findings would be relevant to an understanding of the situation in the home and the school." (Keeves, 1972, p. 282.)

Miller (1970) has surveyed many multivariate studies relating general university achievement to background factors. In addition, a number of other studies have considered the relationship between background factors and achievement in college and first year university science (for example, Bolte, 1966; Even, 1968; Szabo, 1969; Pentony and Loftus, 1970). While many applied system studies have lacked consideration of the basic aspects of the systems approach, many of the multivariate studies have lacked involvement with the developmental aspects of the system being investigated. In most cases no consideration appears to have been given to either the objectives of the courses (even in broad terms) or the components of academic achievement. Multivariate analysis techniques have been used to find the best predictors of "academic success" from a battery of readily available information. Alternatively, criteria tests have been developed and used in preference to teacher examination results. However, the details of the components of such tests and their relationship to the actual classroom teaching have usually been insufficiently clarified.

In an exploratory study of teaching styles in Scottish O-level physics, Pillenger and Houston (1974) did consider the stated aims of the syllabus. They compared groups of students taught by different teaching styles with respect to the attainment of particular goals. The investigation showed that different teaching styles may well effect different goals in different ways. Both aspects of the systems approach are implicitly commented on by Balzer (1970, p. 28) in a review and appraisal of research on classroom behaviour in science. He states;

> "similarities and differences in behaviour of individual teachers should be carefully studied and eventually these data should be related to models and findings concerning effectiveness."

However, in terms of the applied systems approach he states;

> "effectiveness is many faceted and it has become apparent that science educators must come forth with precise statements of desirable objectives ... in any given study effectiveness must be precisely defined ... goals and objectives of the teacher should be clearly incorporated in such definitions."

Kelly investigated the correlations among aspects of achievement in Nuffield A-level biology. These "achievements" were related to the objectives of the course and the results emphasized the need for further research in this area (see Kelly, 1971b, p. 322). Welch (1973) reviewed the research work concerned with the development and analysis of the Harvard Project Physics course in North America. Walberg's work (for example, Anderson and Walberg, 1968; Walberg, 1969a; Walberg, 1970) was particularly orientated to a basic systems viewpoint. It would appear that, to be useful and effective, model building of learning systems should be based on a clear understanding of the purposes of the system and a familiarity with its development.

The overall trends identified in this chapter can be briefly summarized as follows:

(i) Recent trends in curriculum development have been away from large national learning projects to smaller teacher developed learning programmes.

(ii) Trends in curriculum evaluation have been away from emphasis on product evaluation to emphasis on process evaluation.

(iii) Trends in elucidatory research have been toward multivariate studies.

CHAPTER 4

A GENERALIZED CURRICULUM MODEL

> OUTLINE: General diagrammatic models of a first year university physics course are constructed to act as a basis for the development and analysis of the first year physics courses at the University of Waikato. Consideration is given to the statistical techniques available for the development of a specific empirically based model of the learning system. Finally, background information is provided about the first year physics courses.

4.1 The initial model

A model of any system, and how it is explicitly stated (if at all) is a function of a researcher's particular interest. Quite different models will be developed by different people and, as Blalock (1964) emphasized, there is nothing unique or special about any particular model. In terms of curriculum development and teaching, any instructional system can be viewed as an open system with goals where the long-term aim is to identify these goals clearly and to develop ways of achieving them. This curriculum development-teacher orientated view has been used in developing a generalized model (simply a point of view) of an instructional system (Figure 4.1). The "instructional system" is assumed to be a separately assessed unit of work in one subject for one year either at school or university.

From the teacher's point of view the important input variables are the students' knowledge, skills, interests and other background factors which affect student achievement. Although course objectives are determined to some extent from "outside" the system of Figure 4.1, every teacher will make certain decisions to modify the instruction and objectives in terms of the students' background and progress. If all students are basically exposed to the same instructional programme, the particular programme can be assumed to determine the parameters relating to input variables to the students' output achievement. Where a system contains a variety of alternative instructional programmes (e.g., a national system with different teachers, or a course which is taught by different methods to different groups of students) different parameters may apply to the alternatives available.

Any instructional system is part of a larger more complex educational system and any model can emphasize only certain aspects of the interaction of the particular instructional system of interest with this larger educational system. A curriculum developer or teacher would wish to emphasize in such a model those aspects of the wider educational system which appear likely to have some bearing on the decisions which need to be made to improve instruction, or which at least appear pertinent to developing a better understanding of the interaction between course and student. The viewpoint emphasized in the generalized curriculum model proposed in Figure 4.2 is specifically for physics. However it could, with minor modification, apply equally well to other subjects. The emphasis in this diagram on the wider physics education system in New Zealand (longitudinal perspective) rather than on emphasizing the wider educational system of the university (latitudinal perspective) simply reflects the emphasis in this research on one subject area.

These two generalized curriculum models (Figure 4.1 and Figure 4.2) were used to co-ordinate the development and elucidatory aspects of the research detailed in this thesis. Of necessity, these general models are hypothetical in nature, evolved from a review of the research work detailed earlier, from intuition and from discussion. Hopefully both the findings from evaluative studies, and the development of statistical models of specific instructional systems might be useful for the continued development and adaptation of such models.

4.2 Generalized curriculum models and specific statistical models

Generalized curriculum models like those of Section 4.1 are tentative models used to provide a perspective on which to base the development of an instructional system. They also provide a viewpoint from which an empirically based model of a particular instructional system might be evolved.

With respect to the first year physics courses, which are the subject of this study, early consideration was given to the type of models which might be subsequently developed. It was assumed that the structure of the first year physics system would be probablistic rather than mechanistic and that the understanding of the overall system structure would arise from the determination of the statistical correlates between inputs and outputs of the entire system rather than from the determination of individual interactions. If multivariate statistical methods are likely to be useful for model building, the problem was then to determine which of the variety of statistical techniques available would be likely to be most useful and productive. In a review of the literature on the applications of multivariate techniques in educational research, Tatsuoka (1973) indicated that by far the most widely used techniques are multiple regression analysis and factor analysis. Factor analysis can be used to explore the structure of the interrelationship among many variables and as a taxonometric device to develop, or check, measures of theoretical constructs. It can also be used to reduce the number of variables prior to other multivariate analyses. Multiple regression is widely used to develop statistical models. While the necessary linearity assumption is a major one, this simplest assumption seems appropriate in terms of the difficulties in identifying and measuring educational variables. If necessary, variables known to be non-linear may be transformed into new variables to facilitate linearity.

In terms of the generalized curriculum models of Figure 4.1 and Figure 4.2 however, an adaptation of multiple regression analysis called path analysis seemed a most appropriate way of developing a statistical model of first year physics. Path analysis, which is described briefly in Appendix B, has application in a wide variety of fields. Originally developed by Wright (1960a), it has recently been applied to sociology (Duncan, 1966), chemical engineering (Jonnada and Fegley, 1974) and education (Keeves, 1972; Gimmel, 1974).

Again in terms of the generalized curriculum viewpoint of Section 4.1, specific systems models at three levels of complexity can be envisaged, see Figure 4.3. In Figure 4.3(a) (which will be called Model Type I) the elements of the input vector x_i are assumed to be independent variables. While significant correlations between these variables may exist, no causal relationships are assumed to exist. The output variables y_i are assumed to be dependent on the input variables and the parameters of the system T. The matrix relationship $\underline{y} = \underline{\underline{T}}\,\underline{x}$ where $\underline{y}$ = output vector, $\underline{x}$ = input vector and $\underline{\underline{T}}$ = transfer matrix has simplicity and a wide variety of systems can be usefully modelled in terms of this relationship. In an educational setting the elements of vectors x and y may be either directly measured variables or a reduced set of variables obtained from composite scores, e.g., via factor analysis. The elements T_{ij} of the transfer matrix, T, are likely to be the regression coefficients for obtaining the best prediction of y_i from the elements x_i using multiple regression analysis (see Cooley and Lohnes, 1971).

In Figure 4.3(b) (Type II model) causal relationships between variables apart from the input-output relationships may be assumed to occur. For example, output variables may be related to two variables x_1, x_3, however x_3 itself may be dependent on x_1 (e.g., x_3 = school physics achievement, x_1 = earlier mathematics achievement). Path analysis can be applied to this model.

In Figure 4.3(c) (Type III model) feedback has been included. Path analysis may be adapted to include feedback as Wright (1960b) has shown. However, as Duncan (1966) states, it is perhaps questionable whether our present techniques of social science measurement are adequate for the development of such specific statistical models. With respect to the programme modification feedback loop of Figure 4.1, most substantial changes to the course caused by feedback are likely to be of a yearly cycle. Hence, in terms of a statistical input-output model of student skills, knowledge and attitudes over one academic year, this feedback cycle may be justifiably ignored. The student motivational feedback loop of Figure 4.1 however, may be an important aspect of learning which should be included in a model of the system, as this feedback cycle is likely to have a delay time much less than the length of the courses.

Any specific statistical model, while aiming to clarify empirically aspects of the generalized curriculum model, must of necessity be more limited. This is because of the inherent complexity of the system, the problems of identifying and analyzing variables, and the limitations in the statistical models themselves. In the long term the criteria for the success of a model are not that it represents all aspects of reality but rather that it provides useful information for the making of better informed decisions. Indeed, as Blalock (1964) has pointed out, such specific model building in natural settings may at present be of little practical value because the measurement techniques, the control over extraneous variables, and the theoretical tools available may not yet be adequate to the task. Yet, as he states, at some point in time these problems must be identified and faced if we are to analyze and understand non-experimental situations.

4.3 The first year physics courses

The teaching of physics at the University of Waikato was begun along with the teaching of Earth Science, Chemistry and Biological Science in 1970. In that year courses in science were offered at the first year level only. Almost all the academic staff involved in this teaching were appointed only a few months prior to the commencement of teaching. The initial planning and development of the two complementary first year physics courses were primarily student task orientated (i.e., involved with the choice of suitable student activities to achieve implicit generalized course objectives) within the constraints of the proposed degree structure, severely limited staff time, and material resources.

In Appendix A details are provided of;

(i) the New Zealand educational system,
(ii) the degree structure of the University of Waikato,
(iii) the initial objectives and the structural decisions made prior to the commencement of teaching in 1970.

During 1970 this particular project was proposed and from the point of view of a systems approach to the development and analysis of the first year courses, a generalized view of the educational system was diagrammatically developed; detailed course goals were formulated explicitly through staff discussion; the relationship between these goals and the teaching were considered, and methods of assessing the achievement of the goals were investigated. PART B of this thesis considers aspects of this work and the subsequent development of the courses during the ensuing three years.

Simultaneously with the development of the courses consideration was also given to how information might be obtained for the development of a specific statistical model, of the learning system in terms of the course objectives. This aspect of the research study was planned during

1970 and 1971 at a time when it appeared that by 1973 a large number of students ($N > 200$) matriculating directly from school would be attempting the course in each year, and hence a statistically viable population would be available.

Unfortunately by 1973, when the emphasis was shifted from the development of the courses to their analysis, fewer students attempted the courses (due to a change in enrolment patterns in New Zealand universities) than had been anticipated. Data was therefore gathered for two years, 1973 and 1974. After some initial analyses it was found to be more useful to ignore any minor changes in the system from 1973 to 1974, and develop a statistical model from the overall statistics available from these two years. Part C discusses this aspect of the research.

PART B

SYSTEMS DESIGN APPLIED TO A SPECIFIC PROGRAMME

CHAPTER 5

THE DETERMINATION OF OBJECTIVES (CONTEXT EVALUATION)

OUTLINE: The objectives of physics teaching are examined and particular constraints on the choice of objectives for a first year physics course in a New Zealand university are mentioned. Various ways of explicitly stating course objectives are explored and a case is made for generalized learner objectives. The objectives of the University of Waikato first year physics courses, as finally developed, are presented and considered in detail.

5.1 The formulation of the objectives of a physics programme

An awareness of the possible goals of physics teaching, an appreciation of the methods by which different goals might be achieved, and a set of values with respect to the relative importance of different goals in the teaching of the subject seem to be important attributes of physics teachers. Admittedly the set of values needs to be adaptable to a changing environment but it must be sufficiently stable to provide a basis on which to judge the appropriateness of an educational project, textbook, and so on, to the teacher's own developing teaching system. Undoubtedly many very competent teachers have never explicitly formulated the goals they aim to achieve. It is likely that the way they were taught, as pupils, greatly influences the objectives and teaching style of these teachers (Drumheller, 1974).

One of the major problems caused by a lack of explicit consideration of goals is that communication between physics teachers about their particular intentions in teaching and testing becomes difficult. Statements such as "I require my students to have a good understanding of basic physics", may imply rote learning and recall to one teacher, comprehension of the concepts to another, and the ability to apply the knowledge in problem situations to a third. Often teachers feel uneasy about a particular national examination paper without being able to specify why they are not happy about specific questions. Probably the national examiner has, through questions, stressed objectives which the teacher did not consider important because he had not explicitly identified the objectives behind the questions, or because he simply did not appreciate their importance.

The most widely used bases for categorizing educational objectives are those of Bloom (1956) and Krathwohl (1964). Possibly their categories are too generalized to provide an ideal basis for use in a particular subject. However, Bloom *et al.* (1971) have developed categories for various subject areas. In particular, with respect to science, Klopfer (1971), in Bloom *et al.* (1971), has listed a wide range of educational objectives which provide a reasonably broad basis for the explicit formulation of goals for a science course, particularly at the secondary school level.

Ferris has stated that in his view physicists generally seemed to agree on two crucial points:

> "1. Good physics instruction, at whatever level, involves the thorough understanding of a common core of subject matter by no means encyclopaedic in nature, which results in what might be called requisite substantive knowledge.
>
> 2. Good physics instruction equally involves the systematic development of skill in the methodology or process employed by the physicist in investigating the universe around us."
> (Ferris, 1960, p. 270.)

The distinction between these two aspects of physics teaching is also made by Ritchie (1970) who distinguishes between;

(a) teaching the content of physics, and

(b) teaching for an understanding of how a physicist thinks.

Ritchie also details the cognitive objectives stated for physics by the Scottish Certificate of Education Examination Board (1969), Physics O- and H-grades, in terms of Bloom's hierarchy.

In Australia, Mackay (1971a; 1971b) considered a number of objectives for a PSSC type physics course taught in Victorian schools. While these objectives were developed largely from the work of Ferris (1960), Bloom (1956) and Krathwohl (1964), Mackay also included objectives related to cognitive preferences.

In New Zealand, PSSC physics has had a significant influence within the last decade on physics teaching at the secondary school level. PSSC was largely adopted in New Zealand schools at the sixth form level during 1964 to 1968, and then integrated with the "traditional" approach into a new University Entrance physics syllabus largely orientated toward PSSC content and objectives (UGCH, 1970). Many of the objectives stated by Ferris (1960) for the PSSC course are implicit in this sixth form and subsequent seventh form syllabuses (UGCH, 1971) currently in use in schools in New Zealand. These syllabuses (developed by a committee representative of schools, universities, and the Department of Education) do emphasize certain goals but at a level scarcely useful for guiding teaching and testing. However, the Teacher's Guides, published by the Department of Education for each course, tend to co-ordinate and respecify these aims in a manner which is a little more operationally useful (see in particular CDU, 1973).

The choice of objectives for a course is of great importance and, yet, normally there is no clear rationale behind the choice of, or emphasis on, particular objectives in physics. Often, as already implied, individual objectives formulated by others are considered and either adopted or discarded. At the secondary school level, there is a variety of educational programmes (projects) being produced internationally, each project with its own choice of, and emphasis on, goals. Deeson (1974) hints that the lack of a basis upon which to choose objectives, and to determine the emphasis given to objectives, could be a major problem in science education. In an attempt to establish such a basis in terms of the opinion of society, Duncan *et al.* (1971) in New Zealand, sought the views of groups of people in a range of occupations about the proposed objectives of a sixth form chemistry course.[1] These people were chosen because their occupations required a knowledge of, or interest in, chemistry or chemistry teaching. They were asked to rate in order of importance seven stated objectives of sixth form chemistry; knowledge, understanding, skills (of interpretation and communication), application, methods of science (i.e., ability to propose and evaluate methods of solving new chemical problems), philosophy of science, attitudes toward scientific endeavour. While there were distinct differences of opinion among the different occupational groups, the general consensus of opinion was that in chemistry teaching at the sixth form level the objectives should be (in descending order of importance);

(i) an understanding of scientific concepts and relationships,
(ii) the ability to use the methods of science in solving problems,
(iii) the building up of a body of chemical facts and principles, and
(iv) the development of attitudes appropriate to science.

Little weight was given to the other three objectives. Duncan *et al.* (1971) also asked those involved in the study who, in their opinion, would be the most suitable group for the setting of the objectives of a sixth form chemistry course. Although it was widely believed that other groups should have some voice in the matter, it was held that university teachers of chemistry formed the most important decision making group.

The aims of a university teacher are likely to be different from those of a school teacher and this will have an effect on the objectives

---

[1] The groups from whom opinions were sought were: university and sixth form chemistry teachers, scientists, engineers, doctors, dentists, pharmacists and university chemistry students.

chosen for a particular subject at university level. The functions of a university can be divided into three main areas;

(i) to synthesize and disseminate existing knowledge,

(ii) to provide an atmosphere in which people may learn to think critically, creatively and use their initiative,

(iii) to contribute to the growth of knowledge.

The first two functions are undoubtedly of prime consideration with respect to the aims of undergraduate teaching. Often the later years of undergraduate teaching programmes place increasing emphasis on function (ii) with greater emphasis on small group study, reports, individual projects and dissertations. With respect to this function, Beard (1970) states that a group at the University of Surrey considered that the skills which should be acquired from general university teaching included;

(a) the development of articulate expression,

(b) the ability to make independent value judgements,

(c) the development of a capacity for imaginative, creative and abstract thought.

With regard to the views of students, a study by Katz and Katz (1968) at one Australian university indicated that from a wide variety of possible objectives, students considered that the most important purposes of university education was to provide "vocational training and develop skills and techniques applicable to a career". While most university teachers would consider this "training" should be biased towards a generalized education for long term usefulness, there is a strong career orientation in much university teaching.

With respect to first year physics courses in New Zealand a number of additional factors, or constraints, have an influence on the content and objectives of the courses:

(i) The courses have to provide a reasonably easy transition from school to university both with respect to content and objectives. The final year at school for the majority of first year students (seventh form) provides courses very similar to the first year university courses in content, orientation and in at least some objectives.[2]

(ii) The courses must be of reasonable predictive validity so that students are able to assess adequately their likely long term ability in, and attitude towards, that subject. This is particularly necessary at the University of Waikato where students usually study four different subjects in their first year and then "specialize" in subsequent years.[3]

(iii) Apart from the above, first year courses have to be seen as part of a three year course in physics and certain objectives of physics education are considered to be more appropriately emphasized at other levels. For example, experimental techniques and communication skills are emphasized at second and third year levels.[4]

(iv) The courses provide prerequisites for specialist degrees, e.g., engineering, architecture, medicine, surveying, etc. While these specialist schools do not place excessive

---

[2] With university representatives on the writing committees for the seventh form syllabuses and Teacher's Guides, universities in New Zealand have not only an influence on, but a full awareness of, what is taught in schools. In 1970 it was agreed that once a new school syllabus in physics has been accepted by universities and schools, then it is the responsibility of each university to make the necessary adaptations to its own teaching programmes to ease transition from school to university.

[3] Where broad based degrees exist some generosity in grading needs to be shown to students who attempt a subject but find they are not suited to it in terms of ability and/or interest at the university level. The restricted pass system at Waikato University enables a student to count a course towards a degree but gives a clear indication that he should not pursue the subject to a higher level.

[4] Courses 73.204, 73.304, 73.307 and 73.308.

constraints on teaching, "requisite substantive" knowledge for these specialists must be considered. This also applies to chemists, biologists and so on, within the faculty.

5.2 The generality of objectives

The aims or objectives of a course of instruction may be stated in many different ways and at different levels of generality. It is possible to specify them in terms of what activities the teacher expects students to do, or in terms of what learning outcomes are expected (or intended) from a learning programme. The second procedure seems the more desirable as it shifts the emphasis from the learning process to the all important learning outcomes. This emphasis on learning outcomes as Gronlund (1970, p.1) suggests "clarifies the intent of our instruction and sets the stage for an evaluation of that instrument". Eisner (1969a) points out that such formulated behavioural objectives facilitate a number of subsequent functions:

(i) A clear statement of objectives provides goals towards which the curriculum is aimed.

(ii) Once stated they help the selection and organization of teaching and learning activities.

(iii) They provide cues for formulating evaluation procedures.

How such objectives are to be specified, indeed if they need to be explicitly stated at all, depends on the particular type of instructional programme under consideration. For example, a theoretical physics Ph.D. supervisor will explicitly state the problem he expects the student to work on. However, in addition he will implicitly, maybe even unconsciously, expect the student while working alongside him on this problem, to develop certain complex knowledge and skills, ways of thinking and expression, points of view and attitudes, which the supervisor considers important and desirable for a person working in that field. He is likely to consider that education at this level involves such complex and sophisticated objectives that any attempt to state these aims explicitly is of little or no use either to himself, the student, or anyone else. However on the other hand, if an instructor is concerned with "training" someone to operate a machine or carry out routine tasks requiring specific and limited knowledge and skills, specific detailed behavioural objectives are likely to be of considerable benefit in developing an optimum learning programme and pertinent evaluation of learner performance. As Hogben (1970) states, behavioural objectives have been most strongly advocated, and most enthusiastically employed, by "educational technologists" who spend time developing educational materials (e.g., at a national level). For this purpose they need to know precisely what students need to accomplish and what they should be able to do.

Undergraduate university education—as well as secondary school education—lies somewhere between these extremes of educational complexity represented by the "doctoral" and "routine-training" examples given above. Hence the necessity, or at least the desirability, of explicitly stating objectives at either the classroom or national level in these educational areas seems to be worth serious consideration. However, even if it has been decided that explicit behavioural objectives are desirable, what is to be specified and at what level of generality this is to occur, remain to be determined.

Objectives can be written in detailed terms listing the specific behaviour students are to exhibit at the end of the instructional period with different objectives for different content areas. Taken to the extreme this leads to specific behavioural objectives of the type: "Given a simple electric motor, the student must be able to identify, by labelling, at least 10 parts; there will be no penalty for guessing. [list of parts inserted here]." However, if the curriculum is of any substance, or educational complexity, the task of specifying all such objectives becomes formidable. Also flexibility and spontaneity are removed from the course and it is likely that the more complex (higher) educational goals will tend to be ignored. Further, as Wight (1972) suggests, the measurement aspect of specific behavioural objectives of this type may distort the goal.

Alternatively more generalized objectives, or course goals, which are basically task or topic free may be specified; for example, "develop the ability to apply theoretical results to relatively familiar problem situations", "develop the ability to interpret information, particularly mathematical descriptions", "develop the ability to critically evaluate scientific information". In such a case, samples of learning activities (which may hopefully lead to the attainment of such objectives) and samples of learner performance (which we are willing to accept as objective evidence, or indicators, of learning achievement) can be detailed to clarify each objective. These generalized learning objectives become central, and both teaching and learning can be oriented toward these objectives.

Most teachers at all levels in practice tend to be primarily activity or topic oriented. Usually this is because of constraints placed on their teaching by external syllabuses, by the equipment available for experimental work, by the importance in teaching given to knowledge objectives, or by the practical problems of classroom management. However, if generalized learner objectives exist, topics may be adapted or changed to optimize the achievement of all objectives. For example, a particular generalized learner objective might be; "students should develop an awareness of the limitations and approximations in physical theories". A teacher can then consider how he can;

(i) adapt the activities within a particular content area to develop this objective,

(ii) use that content area to develop measures of the attainment of this objective.

Certain activities and content areas are more fruitful than others for teaching and/or evaluating performance of a particular objective and a teacher, or curriculum planner, needs to be vigilant for the opportunities that may be available within the constraints of the system. Campbell (1972) points out that in a task or topic oriented approach, an awareness of objectives can help guide the teaching in the shaping of approaches and methods, and can give a long-term purpose to the task, or topic. Generalized learner objectives, which will be called course objectives in this study, then become pervasive goals that remain constant while tasks and topics change.

Assuming that specific behavioural objectives related to particular tasks or topics are not desirable, at least in the first instance, course objectives can be developed and clarified by considering how proposed objectives would relate to both learning and assessment procedures.[5]

To decide just how to state the course objectives it is desirable to consider the possible purposes and the potential usefulness of such objectives. It would appear likely, that provided the objectives are stated in a suitable way, they;

(i) could be useful for teachers as a guide to orientate and diversify teaching, to organize and diversify testing, and to communicate with fellow teachers,

(ii) could be useful for students, along with examples of

---

[5] For certain methods of classroom teaching and testing (e.g., individualized instruction, criterion-based feedback) detailed specification of what the student should know with respect to a task may have a place. However, these specifications usually cover only minimal requirements and relate to "low level" objectives and content coverage. Provided course objectives are clear and these detailed specifications are seen as relating to a limited number of the objectives only (hence not taken as <u>the</u> course objectives) then such specificity may be useful in certain instructional systems.

expected learner performance, as a guide to learning, not as hurdles to be overcome but as a means of orienting themselves towards the goals of the course,

(iii) could be organized into a logical and philosophically satisfying sequence and hence provide a unity of purpose to the whole programme,

(iv) could either be the goals of the course, or relate directly to the goals, hence the overall aims of the course could be continually kept in mind,

(v) could be sufficiently few that they could be transferred from a set of ideals listed on paper to an operationally active viewpoint in the mind of the curriculum planner or classroom teacher,

(vi) could be sufficiently general so that different physics programmes would be likely to have some common as well as different objectives and hence similarities and differences between various courses could be appreciated.

In terms of the purposes outlined above it would seem likely that the level of generality at which objectives should be stated depends on the particular person who will be using the objectives. For a teacher, or student, new to the concept of learner objectives for a course, five or six objectives which emphasize the major objective areas may be all that can be handled. For a curriculum planner or examiner such generalized objectives would be inadequate. He would be aware that in terms of teaching and/or testing there are a number of distinct aspects within a general objective and possibly, that there are also additional objectives not in his original framework of objectives which need to be included. However, even at this specialist level it appears unlikely that more than 15 to 25 objectives will be necessary to cover the knowledge, skills, and attitudes which are considered important in a particular course, or to be useful in terms of the purposes previously outlined for such objectives. Provided the list of objectives covers all aspects of the knowledge, skills and attitudes appropriate to the course and tend to diversify rather than restrict teaching and testing, further subdivision of objectives is likely to confuse rather than assist curriculum planners or examiners, particularly in their communications with others. Care must be taken to ensure objectives do not become so extensive that they overwhelm or confuse (Duchastel and Merrill, 1973). Finally, not only the way the course objectives are specified, but also the particular samples of learning activities, and/or learner performance criteria, must be appropriate to the experience and knowledge of the person to whom the objectives are being communicated. In particular, considerable respecification of objectives and sample behaviours are necessary for students, if they are to clearly understand and appreciate the course objectives.

5.3 Objectives of the first year physics courses

As detailed in Appendix A, the two physics courses offered at first year level at the University of Waikato (73.101 and 73.102) are complementary and nearly all students take both courses. Basically, during the period of this research, the two courses together covered the content of the first year physics courses as then offered at other New Zealand universities. Although the two courses were separately assessed, the distinction between them was primarily in terms of teaching methods. In many other ways the two courses formed a single entity. Initially lists of course objectives were developed for each course separately; however it was subsequently found to be more useful to develop objectives for the two courses together. Because each course was taught by different teaching methods, some of the objectives were more appropriate to one course rather than the other, while other objectives were relevant to both courses. (This will be considered in

Chapter 6.)

A tentative set of explicit objectives for the first year courses was developed in 1970 from implicit objectives (Appendix A.3), in the light of the literature surveyed in Section 5.1, and in terms of the philosophy of Section 5.2. This tentative set of objectives was used to guide the initial development of the programme in 1970 and 1971. The objectives were adapted and added to (particularly with respect to higher skills) after further consideration by the other academic staff involved in the courses. To some extent the work of Klopfer (1971) was used to rationalize these objectives, and to order them in a philosophically satisfying and practically useful manner. Feedback from teaching (Chapter 7) and student evaluation (Chapter 8) helped considerably in the further development and clarification of different aspects of the objectives of the courses. Increasingly the objectives were used when the teaching activities of the courses and the assessment details were being reviewed. At the same time, the objectives themselves have been kept under review.

The objectives developed by 1973 can be grouped under seven major headings, namely, Knowledge and Comprehension, Application, Analysis and Higher Skills, Processes of Experimental Analysis I (Experimental Design), Processes of Experimental Analysis II (Analysis and Conclusions), Orientation (Nature of Physics) and Attitudes.

The objectives are specified in Table 5.1 and explained in greater detail below:

I Knowledge and Comprehension of Physics

Emphasis is placed in the first year courses on the comprehension and understanding of the basic principles of classical physics, elementary quantum ideas and simple relativistic concepts. This content was considered both appropriate for students living in a swiftly changing technological world and a suitable basis for specialist studies not only in physics but also in engineering, architecture and so on. Chapters 1-12, 15-17, 19-21, 23, 24, 26-29 (with parts of Chapters 18, 25 and 30) of Physics by Alonso and Finn (1970) covered the content at a suitable depth of treatment.[6] The unambiguous specification of content based on the textbook was considered desirable. The understanding of fundamental principles was considered more important than the memorization of specific details. This set of objectives is therefore not concerned with recall but with the ability to acquire and comprehend information and to understand basic physics. Little distinction was seen between "knowing", "understanding" and "being able to apply knowledge when specifically asked to do so".

The three course objectives grouped under the Knowledge and Comprehension heading were:

I(A) Develop the ability to extract information (from scientific literature)

This objective is included primarily to encourage students to learn to use reference material. Students instructed by a lecture notes method prior to entering university often do not think of using reference material, even their own textbook, for the purpose of acquiring requisite information.

I(B) Develop the ability to interpret information, particularly to interpret mathematical and graphical descriptions

This objective is one level above I(A) in that its achievement requires the additional ability to comprehend information, and to translate it from one form to another, e.g., graphical, mathematical or verbal.[7] Naturally this often requires previous comprehension of

---

[6] The only major topics of classical physics omitted were statistical mechanics and thermodynamics. These topics were considered more appropriate for a second year course.

[7] The interpretation of experimental results via graphical analysis was considered under Objective IV(E).

earlier material. Emphasis at this level is particularly on the ability to comprehend ideas expressed mathematically, e.g., that $y = A\sin 2\pi(\frac{x}{\lambda} - \frac{t}{T})$ represents a travelling wave (see Osborne, 1972).

I(C) Obtain an understanding of the facts, theories and principles of physics (as specified in the courses)

Understanding here implies knowledge and comprehension and includes comprehension of how a specified concept or principle applies to at least some particular situations. For example, the concept of Gauss' Law is not useful unless the knowledge acquired includes how it might be applied to at least a few specific situations. "Specified in the course" implies that the objective is concerned with content that is clearly understood by the student to be part of the course. (This content might be specified, and covered, in lectures, or students might be directed to it as part of their laboratory activities.) Knowledge and comprehension of content not directly taught, for example, acquired by extrapolation, awareness of implications, or pure interest is assumed to involve higher intellectual skills or attitudes, and consequently is not included as part of this objective.

II Application

The application of facts, theories, and principles to solve problems involves an intuitive judgement about which equation or principle is the most appropriate to apply to the problem situation. The problems set do not specify how they are to be solved. If a problem specifies "use Gauss' Law to ...", it is essentially a question investigating a student's understanding (Objective I(C)) of Gauss' Law. However, if the problem does not specify Gauss' Law then it not only requires a knowledge of Gauss' Law but an intuitive judgement that the problem can be solved by Gauss' Law. Problem solving also requires confidence in one's ability to obtain a solution—successful problem solving encourages successful problem solving.

Emphasis on routine problem solving has decreased during the last ten years and this is unfortunate in that routine problem solving probably provides a sound basis for more complex problem solving, as well as providing opportunities for students to become competent at basic mathematical skills.[8]

With regard to these courses three levels of application, or problem solving, were identified and emphasized.

II(A) Develop the ability to apply theoretical results to relatively familiar situations

The simplest problems are those which require the recall, or extraction, of an equation and the substitution of numbers into that equation to yield a solution. Some physics teachers consider that this "recall and substitution" is not "application" at all—that this ability is not a worthy physics objective. However, many students have difficulty in deciding what law or relationship to apply to even the simplest, or most obvious, problem situation if the law or relationship which needs to be recalled is not specifically stated. These "recall and substitution" questions have been classified as problems which require this lowest level of application ability. One reason for discrediting this type of problem and the ability to solve it has been that, provided the correct equation can be recalled, problems can often be solved without a real comprehension of the physics involved. Obviously, if this objective were overstressed it might well have a detrimental effect on overall learning. However, as an encouragement to more complex problem solving, and as a simple ability useful to a physicist or engineer who needs to be able to recall, or find the appropriate equation and to correctly perform basic mathematical

---

[8] Recent moves toward a more technological physics (see Woolenough, 1974) is possibly a result of an overemphasis in some physics teaching on comprehension without sufficient consideration of how this knowledge is useful to solve problems, particularly real-to-life problems.

manipulative skills, this objective cannot be ignored.

II(B) Develop the ability to apply theory, and mathematically analyze, problem situations

Common to physics is the problem situation where an unknown relationship between quantities has to be derived from known basic relationships which apply to the problem situation, but which themselves are not specified. Often the problem situation must be analyzed to determine the basic relationships which apply. The solution is then derived from these. The level of application is such that for a successful solution a comprehension of the requisite physics is normally required.

II(C) Develop the ability to apply theory to relatively unfamiliar problem situations

The highest level of application involves the ability to apply a principle or law of physics in a novel situation where it is not immediately obvious what physics to apply. While problems classified in II(A) and II(B) require some form of intuitive judgement, difficulties in obtaining a successful solution also arise because of the need to ensure that the relationship is correctly expressed, that substitutions of algebraic or numerical data are correct, and that the correct mathematical computations are made. However, in this II(C) objective, problems are characterized by the fact that the key difficulty is the "finding" of the particular principle or law which will enable a successful solution.

III Analysis (and higher skills)

While certain information is specifically taught to students, some of the objectives of a university course should be oriented towards stimulating more complex intellectual skills such as independent, critical and creative thought. For example, stimulation may be provided by the discussion of ambiguous or paradoxical situations, by encouraging students to question the assumptions and implications behind the prescribed knowledge, and by expecting students to synthesize knowledge into a coherent and logical understanding of physics. While these three aspects of "higher skills" are certainly not all-encompassing they were considered appropriate for this level and were isolated as course objectives.

III(A) Develop the ability to critically analyze problem situations and evaluate information

Much intellectual activity involves criticising one's own work, or the work of others—of isolating mistakes, of identifying inconsistencies, and of uncovering the illogical. This objective is concerned with encouraging this activity in the hope of developing a critical, yet constructive attitude. Particularly at this level emphasis needs to be given to the criticism of textbook-like material because students at this level are most uncritical of published works.

III(B) Develop a critical attitude towards, and an ability to analyze, idealizations, approximations, and assumptions in physical theories

While knowledge of relationships is important, and an ability to recall and apply them in simple problem situations is not without merit, a clear appreciation of the limitations, approximations and idealizations in a relationship is necessary to ensure that the relationship is really appropriate for the situation under consideration. While many will argue that the knowledge of limitations and approximations in particular situations is inherent in understanding physical relationships (Objective I(C)), an overstressing of this aspect when ideas are initially presented appears to lead to confusion at this level (Frisch, 1974). For this reason, although to some extent limitations and approximations are necessarily presented in lectures, emphasis should rather be on encouraging students to appreciate the importance and implications of such approximations and, once they have comprehended the basic ideas, to be critically aware of the limitations imposed on derived relationships.

III(C) Develop an understanding, and appreciation, of some of the major themes and unifying concepts of physics (including the relationship between fundamental principles and derived laws)

It may well be argued that this is merely a Knowledge and Comprehension objective, like Objective III(B). However, it was grouped as a higher skill because of its place in the development of understanding.

Only after acquiring substantive knowledge can the major themes and unifying concepts of the subject be appreciated, and the complex relationships between "fundamental" and "derived" laws clarified. To gain a deep understanding of physics students need to develop or synthesize a personalized theoretical framework into which future learning can be fitted. This synthesis requires the student to consider thoughtfully the relationships between recently acquired knowledge and earlier learning and to eliminate apparent inconsistencies.

IV Experimental Skills

A number of intellectual skills associated with experimental physics have already been considered, in that experimental physics, somewhat unlike practical physics, is basically concerned with the relationships between reality and theory. Hence Objective III(B) is closely associated with experimental physics, as are the knowledge and comprehension objectives. Consideration was given to the desirability of integrating the experimental objectives into groups I, II and III above. However, from the point of view of teaching and testing it was found more useful to consider them separately, as indeed Klopfer (1971) and Ritchie (1970) have suggested.

Two aspects of experimental work were considered important in the experimental course (73.102)—first, the planning of experiments, and the abilities associated with this aspect of experimental work, and secondly, the recording and subsequent analysis of experimental results and the deduction of conclusions. There are, of course, other aspects of experimental work—for example, the selection of measuring instruments, skills in using laboratory equipment, and the performance of laboratory techniques. However, it was decided (see Appendix A.3) that the acquisition of practical skills, as such, should not be stressed at this level, and that the main emphasis should be on understanding physics and developing intellectual skills. This emphasis in the first year course is complemented by a second year physics course which emphasizes experimental techniques.

IV(A) to IV(C) Processes of Experimental Inquiry I (Experimental Design)

IV(A) Develop the ability to propose tentative ideas, and working hypotheses, and make logical and testable predictions

The ability to propose working hypotheses and tentative ideas is required both in experimental physics and in applied technology. In experimental physics there is a need to be able to propose a variety of tentative hypotheses when theoretical results do not agree with experimental results. In applied technology a variety of tentative ideas based on known physical principles or phenomena may be required before a practical solution to a technological problem is found. Apart from requiring divergent rather than convergent thinking, the search for physical phenomena to account for experimental results, or to solve a practical problem, is similar in many respects to the search for the physical principles required to solve theoretical problems (Objective II). As in Objective II, a breadth of knowledge and an intuitive judgement are required if the problem is to be solved. However, in physics pure novelty is insufficient, the tentative ideas must be reasonable and useful, and lead to testable solutions.

IV(B) Develop the ability to translate problems into experimental operations

Every experimental investigation involves the translation of a problem into experimental operations. (How should we go about testing this idea, or investigating this phenomena with the equipment available?) This objective is concerned with developing a student's ability to plan, in general terms, a workable experiment.

IV(C) Develop the ability to make useful and reliable observations

To be able to plan an experiment in general terms is not enough. One also has to decide what measurements should be taken and what degree of accuracy is required. For example, if a relationship is non-linear, should equally spaced values be chosen for the independent variable? This objective is concerned with the actual use of apparatus by the student. In a more practically oriented course it would form part of a major group of objectives which emphasized the selection of appropriate measuring instruments, the skills required to use particular equipment, and the actual manual skills required to perform common laboratory techniques such as the wiring of circuits, taking photographs, and aligning optical systems.[9]

IV(D) to IV(F) Processes of Experimental Inquiry II (Analysis and Conclusions)

IV(D) Develop the ability to keep adequately detailed and organized records of experimental work and to be able to extract information from these accounts

Experimentalists need to keep adequate records of their results. These records are not like the neatly-written-up accounts so often expected of students in practical courses, but rather they will be on-the-spot accounts packed with details, data, comments, analyses and conclusions. Such accounts provide information readily available for writing a report or paper at some subsequent point in time. This objective places emphasis on developing in students the ability to keep adequately detailed records of experimental work. Records from which students can subsequently extract required information.

---

[9] In fact, because the experimental course was firmly based on laboratory work, these latter skills did develop quite naturally during the year.

Because of the course policy not to concentrate on practical work, the ability to synthesize material into succinct reports was not an explicit objective of this course.[10] However, the ability to write reports could well be an objective of an experimental physics programme. The absence of this objective in the first year physics courses was offset by emphasis on this objective in second and third year courses (73.204, 73.304, 73.307).

IV(E) Develop the ability to analyze experimental data mathematically and graphically

Physics requires considerable analysis of experimental data, both mathematically and graphically, in order to reduce the data to a form from which conclusions can be deduced. This objective involves the ability to compute an error analysis and to analyze data graphically (including power and exponential relationships).

IV(F) Develop the ability to draw valid conclusions from observations and data

This objective is closely related to a clear analysis of experimental data. Students should be able to decide whether or not the results confirm a tentative idea or a physical relationship. If a systematic error has occurred, it may be possible, in the conclusion, to account for this discrepancy in terms of some obvious inadequacy apparent in hindsight. However, if the cause of the systematic error is not obvious, the ability to frame tentative proposals to account for this discrepancy is not included under this objective because it is included more appropriately under Objective IV(A).

V Orientation

Orientation is primarily concerned with student awareness of the nature of physics and science in order that students may recognise the

---

[10] However it was required to some extent in examination and test questions related to Objectives IV(D) and IV(F).

the place of science in the total field of human endeavour and in other ways of thinking. While there are other aspects of orientation such as the realization of the relationships between scientific progress, technological achievement and economic development and an awareness of the social and moral implications of scientific inquiry, these were not considered appropriate objectives for the courses under consideration. However, certain philosophical aspects of science seemed to be appropriate, particularly as they appeared related to other objectives of the course. For example, the synthesis of a theoretical framework (Objective III(C)—what are the basic principles?), experimental inquiry (Objective IV), and attitudes toward physics (Objective VI) were all considered to require some appreciation of the nature of the subject.

The objectives which follow might well have been considered as part of these other objectives. However, it was decided to view orientation as a separate group of objectives covering aspects of both the cognitive and attitudinal domains (Krathwohl, 1964).

V(A) Develop an awareness of the function and limitation of physical laws

While this objective is closely related to Objectives III(B) and III(C), it is more concerned with developing a generalized viewpoint and less concerned with details of specific situations. Students even at this level, need to be aware of the fact that the laws of science are not proven "indestructibles" but are rather useful summaries of the physicists' view of the world.

V(B) Develop an awareness of the relationship between experimental observation and theory

Many students view an unaccounted-for discrepancy between theory and their own experimental results as simply due to "inadequate equipment". This kind of thinking is liable to influence in quite the wrong way their view of the relationship between experiment and theory in physics. Hopefully, students will learn to appreciate what is implied by Joliot-Curie's remark, "The further an experiment is from theory the closer it is to the Nobel Prize".

V(C) Develop an awareness of physics as a creative, open-ended and continuing development, both tentative and unfinished

Because this objective is concerned with the nature of physics it has been included as an orientation objective. However, this objective is more attitudinal in nature than the other two and could well have been categorized as an attitudinal objective. An awareness of physics as an on-going process is the first step in a student's commitment to an intended career in physics.

VI Attitudes

Attitudes towards any subject are not easily isolated, defined or categorized, and aspects of attitudes are inherently contained within all the previously considered cognitive objectives. Often certain aspects of scientific work are included as attitudinal objectives in a science course. However, as Klopfer (1971) suggests, attitudes such as open-mindedness, self-criticism, willingness to suspend judgement, and commitment to accuracy are really a reflection of the nature of scientific inquiry. For this reason, in these courses, such attitudes are assumed to be inherent in some of the higher level cognitive objectives (particularly Objectives III(A) and (B), IV(A), (C) and (F)) and have therefore not been included as separate attitudinal objectives. Moreover with respect to both teaching and assessment, such attitudes certainly form part of particular cognitive objectives.

Consequently the attitudinal objectives have been limited to those objectives which are characterized by an emotional component (see Krathwohl (1964) categories 2.3 to 3.2) which must be a purely voluntary response on the part of the student. In terms of teaching and assessment they are characteristically different from cognitive objectives where imposed standards and pressures occur. Three objectives were considered to warrant particular attention.

VI(A) Develop the attitude that mathematics is a useful, convenient and precise language in physics, and that the use of mathematics in physics is a satisfying activity

An initial implicit objective was that the theoretical course in particular would concentrate on an overview of the major principles of classical physics at a "reasonably sophisticated mathematical level". Mathematics is a convenient, and precise language in physics and a useful tool. However, like "too-much-Shakespeare-too-soon" in an English programme, there is a danger that students may become negative in their attitude to mathematics if the level is too sophisticated. For this reason Objective VI(A) was isolated as one of importance.

VI(B) Develop the attitude that physics is an important, relevant and satisfying activity

A student's attitude to his immediate activities in physics may be different from his attitude towards the activities of physicists. However, a student's attitude to his learning activities in physics is important and is likely to influence his developing attitude to the activities of physicists. Hence the objective relates closely to a student's experiences and the activities associated with the course. As Klopfer (1971, p. 578) mentions, "there is strong psychological evidence that students learn better, learn more, and remember longer when they find pleasure in the learning experience".

VI(C) Develop a commitment to actively searching for an understanding of physical phenomena and for the theories by which they may be explained

This objective is concerned with interesting students in physics so that they will pursue their interest in the subject beyond the required activities. While most students have chosen their careers before reaching university (and all students in these courses are interested in a science-related occupation) it is also hoped that the uncommitted student may be encouraged to take up a career in physics or a related activity.

Other Objectives

Objectives other than those of the immediate subject area sometimes need to be included. Because of the mathematical level required in these courses a level of competence was required in mathematics, particularly algebraic manipulation, calculus, vector algebra, and numerical computation. While much of this work was covered in supporting mathematics courses, certain aspects of this work needed attention in the theoretical course. The abilities required here corresponded closely to those required for Objective II(A) above, at a more routine level, but in the field of mathematics. Wilson (1971), in developing a taxonomy of educational objectives for mathematics, classified the ability to solve routine problems as the lowest level of his application grouping.

> "Essentially the student is asked to perform a sequence of comprehension-level behaviours and carry out an algorithm to arrive at a solution. For this ability the prime concern is solving the routine problem."

For the theoretical course one additional objective was added.

VII Mathematical Physics

VII(A) Develop the ability to manipulate and use the mathematical language of physics

This objective is particularly concerned with vector algebra, line integrals, and surface integrals. Also a competence to manipulate algebraic and numerical expressions and to handle elementary calculus is required.

5.4 The compatibility of the objectives with other courses

The objectives of a course cannot really be usefully considered in isolation from the teaching and assessment of the course and these two aspects will be considered in detail in Chapters 6 and 8. As implied earlier, the objectives summarized in Table 5.1 were certainly not just a list of objectives extracted from a variety of educational programmes. The original list of tentative objectives was modified and refined by the staff members involved in the courses. The objectives were used as a guide for planning learner activities and for assessment for four years, with further adaptation as and when required to make the order and meaning of these objectives more logical and more appropriate to this level of education, and hence more useful as a guide in the development of the educational programme. In addition, the courses were evaluated by a very experienced secondary school physics teacher, while holding a fellowship at the University of Waikato for 1973. During this time he was able not only to study and comment on the appropriateness of the objectives for a post-secondary school course, but also to study in detail the relationship between the objectives and the teaching and assessment in the courses. Well informed external criticism of this type (he attended all lectures, supervised tutorials and laboratories, and studied all assessment procedures in great detail), albeit only subjectively, is of tremendous benefit in all aspects of course evaluation.[11]

The determination of the objectives of a university course are normally the sole concern of the department offering the course, and can only be developed and clarified by considering the objectives along with the teaching and assessment procedures. However, as stated earlier, the compatibility of the objectives with any preceding and subsequent physics courses taken by students, is desirable. From this point of view one may evaluate the explicitly stated objectives, in isolation from the teaching and assessment of the course, as one aspect

---

[11] It is unfortunate that the opportunity for such study and comment is very rare in most New Zealand universities. Only one teaching fellow is offered per university per year at present. On this basis each university department at Waikato can expect a fellow just once every two decades.

of context evaluation.

Naturally as the subsequent physics courses at Waikato were developed within the same small department as the first year courses it was possible to ensure that the objectives of the first year courses were compatible with the objectives of the second year courses.

Subsequent to the development of the objectives of the first year physics courses at Waikato, the objectives of the national seventh form programme were clarified and ordered into four major groupings (see CDU, 1973, p. (iv) and (v)). These major groupings corresponded very closely to the first five complexes of objectives(I, II, III, and IV (A to F)) of the first year courses. The seventh form Physics Teacher's Guide (CDU, 1973) also mentions the importance of attitudinal objectives.

Because of the basic similarity between the seventh form physics objectives and the first year physics objectives the two sets of objectives were considered to be most compatible. However, the opportunity was taken (in conjunction with some related work in Section 13.2) to assess teachers' views on the importance, in teaching, of those "complexes" of physics objectives which were basically similar in both sets of objectives (the "complexes" I, II, III, IV and VI of Table 5.1). The views of seventh form teachers contributing students to the first year physics courses at the University of Waikato were sought.

Teachers were asked (see Appendix N):

"What importance do you consider should be given to the following physics objectives in teaching at

(A) seventh form level

(B) first year university?

P1 Comprehension of basic concepts.

P2 Problem solving ability.

P3 Critical evaluation of ideas.

P4 Experimental procedure skills.

P5 Processing of experimental data.

P6 Attitude towards physics."

An additional information sheet was provided, explaining each of these aspects in some detail (see Appendix N—Knowledge and Abilities in Physics). Teachers were asked to rank each of these in terms of Very Important, Important, Possibly Important, Not Important, Undesirable.

Of the 41 teachers involved, replies were obtained from 32. Of these, two refrained from commenting on university teaching. No teacher considered any objective undesirable, and the percentages of replying teachers who chose each "importance" category, for each "objective", are given in Figure 5.1.

A mean value of importance was obtained for each "objective" from the replies, using an "importance" scale ranging from Undesirable = 0 to Very Important = 4. These mean values (averaged over all replies) are shown in Figure 5.2.

Four points seem worthy of comment:

(i) All objectives were considered important, mean values almost invariably being between important and very important.

(ii) Basic comprehension was considered of prime importance both at school and at university. Possibly this is because of the apparent hierarchical nature of the objectives.

(iii) Critical evaluation of ideas was considered distinctly more appropriate for university level learning than for school learning. One teacher stated that he considered this objective involved a maturity of thought and approach which many seventh formers do not have because of physical age and "time of contact with the concepts involved in physics".

(iv) There was considerable divergence of opinion over the importance of attitudes. However, this reflects teachers' opinions about their relative importance in teaching rather than their absolute importance. Some teachers who scored this objective low stated that their view was, "teach them physics and let the attitudes look after themselves".

Not only were teachers given the opportunity to add qualifying remarks, but they were also given the opportunity to add objectives which they considered were important. Three teachers, while not disagreeing with the objectives stated, considered that the seventh form syllabus should be more applied or technologically oriented.

Only one teacher wished to add further objectives. He considered that the objectives set out covered all that is desirable except perhaps

"(i) an appreciation of the philosophy and nature of physics, and the position of physics as an integral part of science,

(ii) an ability in the use of mathematics as a language, within the bounds of the syllabus."

It is interesting to note that although this teacher did not have recourse to a list of first year physics objectives, the two objectives he identifies are very closely related to the two "complexes" of objectives of the first year physics courses (Objectives V and VII) not covered in this analysis.

CHAPTER 6

THE DESIGN OF INSTRUCTION (INPUT EVALUATION)

OUTLINE: It is pointed out that one of the major problems of instructional design is to determine the best teaching method to achieve any particular objective. As the first year physics courses developed, and objectives were clarified, various teaching methods were seen as having particular potential to develop certain objectives. The final design, in terms of the relationships between the objectives and the teaching, and between the objectives and the assessment procedures, is discussed in some detail.

6.1 Introduction

Instructional design needs to take into account not only specific objectives but also the teaching techniques which appear to have potential for the achievement of these particular objectives. Input evaluation is concerned with such questions as: What methods or strategies already exist with the potential to realise the objectives? Is it logical to believe that a given strategy will develop a particular objective? A major problem of instructional design is how to obtain adequate answers to this type of question. At present in designing instruction any input evaluation which is undertaken is rather *ad hoc*, as Stufflebeam *et al.* (1971) has pointed out. Decisions on proposed instructional design tend, of necessity, to be based on subjective judgements about learning systems similar to the one proposed.

Undoubtedly there are considerable constraints on what teaching methods can be used in a particular teaching programme. However, provided there is a variety of teaching methods and explicitly stated objectives, there are usually considerable possibilities for adapting the teaching in order to achieve more effectively the course objectives within the constraints of timetabling, room allocation, staff/student ratios and so on.

Consideration of the first year physics teaching in terms of the explicitly stated objectives tended to diversify, rather than restrict, teaching approaches and learning activities. In Section 6.2 the relationship which developed between the objectives and the various teaching methods used in the courses will be outlined. In Appendix E, the relationship between each objective and the teaching and learning activities of the course is discussed in greater detail, for reference purposes.

6.2 Teaching and the objectives

Five distinctly different teaching and learning activities operated from 1971, for the two first year physics courses. In order, from the most expository to the most heuristic, these were:

	Course	Frequency	Staff:Student Ratio	Average Time/Week
(a) Lectures	73.101	2 lectures per week	1:100+	2 x 50 min.
(b) Evaluation session	73.101	1 session per week	1:100+	1 x 50 min.
(c) Tutorial session	(mainly) 73.101	1 tutorial per week	1 : 16	1 x 50 min.
(d) Laboratory session	73.102	1 six hour session per fortnight	2 : 32	3 hrs. (av.)
(e) Independent study	73.101/2	2 to 3 hours per course per week	-	5 hrs. (av.)
				12 hrs.

In all, students were expected to spend a total of 6 hours on each course for each of the 26 teaching weeks in the year. Of these 6 hours approximately 50% was spent in formal contact with academic staff.

The different learning activities, described briefly below, had various focii with respect to the different course objectives.

6.2.1 Lectures

The lectures were primarily concerned with knowledge and comprehension (Objective I). There has been much criticism of the lecture as a method of teaching, but as Beard (1970, p. 91) states,

> "Students praise lectures which are clear, orderly synopses in which basic principles are emphasized ... students of science subjects consider that a lecture is a good way to introduce a new subject ...".

The lecture is an economical way to disseminate knowledge, and the first year physics lectures were primarily to help students to locate, comprehend, and understand the physics detailed in a textbook of a suitable difficulty, emphasis and content. This intent was supported by open textbook examinations and in this way it was hoped that the over emphasis likely to be given to lecture notes (see Beard, 1970, p. 92) would not occur. After the first year (in which students were not encouraged to take notes) lectures were given which did not discourage students from taking notes because students were found to be incorrigibly conditioned to note taking (see Section 7.2). Lecturers used blackboard, duplicated notes, demonstrations, overhead projector transparencies (mainly for diagrams) and Super 8 film loops in their teaching.

Opportunities also arose in the lectures to consider aspects of Objective II, apart from Objective I, particularly through worked examples. Also Objectives III(B) and III(C) were stressed by emphasizing limitations, approximations in theories, and providing comments which might assist students to clarify in their own minds the major relationships between principles.

6.2.2 Self-evaluation and revision session

Many educationists have emphasized the importance of feedback in learning (e.g., Beard and Bligh, 1971). Although initially three lectures per week were given, after the first year (1970) the third lecture was replaced by a self evaluation and revision session which was seen as a possible way of effecting recapitulation and feedback. The policy was that no new information would be introduced during these sessions but rather:

(i) A summary would be provided of the work covered in the previous lectures (see Appendix C.1 for an example. This example is perhaps rather more mathematically oriented than most).

(ii) A short multiple choice "test" would be provided. Students would mark their own work during the session for their own benefit. No marks would be recorded. (See Appendix C.2 for a typical example. Problem sheet questions 1-4.)

(iii) Ideas presented in the lectures would be applied through a consideration of further problems or worked examples (see Appendix C.2. Problem sheet questions 5 and 6).

(iv) Comment would be made on applied aspects of the lecture material. Further, appropriate film loops and demonstrations reinforced ideas already presented in lectures.

These evaluation sessions were therefore primarily concerned with Objectives I and II.

6.2.3 Tutorials

While the lectures and evaluation sessions were basically expository in approach, involving a large group of students (all students together $N \simeq 110$), tutorials provided a small group situation ($N \simeq 12$) where students did have an opportunity to direct their own learning to some extent.

Initially tutorials were concerned primarily with Objective II, problem solving, and the re-emphasis of lecture points, where necessary

(Objective I), to provide sufficient comprehension for successful problem solving. However, as the programme developed, it was found that tutorials provided, through discussion, an excellent opportunity to encourage students to think critically (Objective III(A)). Tutorial sheets were provided at each tutorial session (a) to list problems previously set in lectures and to provide correct answers; (b) to provide questions to start, and focus, discussion (see Appendix C.3).

6.2.4 Experimental work

The experimental sessions undoubtedly provided the richest opportunity for covering a wide variety of educational objectives. Students normally worked together in mixed ability pairs.[1] Each 6-hour experiment provided a set of objectives based on a particular physical phenomenon (e.g., ultrasonics, cathode raytube, microwaves) or a particular topic (e.g., experimental errors, two-dimensional dynamics, force and energy). To guide students a laboratory guide, Experimental Physics, was written for the experimental course (see Osborne, 1974a, also Appendix D). This guide was continually adapted and changed each year (as were some of the experiments themselves) in the light of experience and student feedback in order to ensure that the learning experiences were as appropriate as possible to the objectives of the course.[2] The experimental sessions, and the associated follow-up private study, provided opportunities to emphasize;

(i) all aspects of Objectives I and II,

---

[1] Students were paired primarily because of practical constraints but there appeared to be some advantage in mixed ability pairing. The little evidence available (Hartley, 1973) suggests it is unlikely to be disadvantageous to learning. (See also Appendix E, p.316.)

[2] In developing Experimental Physics, the course objectives were continually kept in mind and opportunities to achieve particular objectives continually looked for. Although this would not normally be done, the text for Experiment 11 in Appendix D.3 has been annotated to show the perceived relationship between activities and objectives.

(ii) Objectives III(B) and V,

(iii) Objectives IV(A) to (C),

(iv) Objectives IV(D) to (F).

Also this work was closely associated with the development of attitudes. Students were clearly informed of these objectives, in terms appropriate to their level of understanding (Appendix D.2) and the relationship between the objectives and the assessment procedures was specified with examples (Appendix D.5).

6.2.5 Independent study

Independent study (individual or informal group study) was primarily directed and encouraged through implicit rather than explicit pressures on students. Problems (Objective II) were specified in lectures and laboratory work, and students were encouraged to attempt these and ask in tutorials for assistance with the problems they found difficult. Also experimental analyses were completed, theory was related to experiment, and conclusions were written up, outside the formal class contact periods.

Normally a suitable environment within the university where first year students can study outside the laboratory, lecture, or tutorial sessions is provided only in the university library. However, as students are expected to spend about 50% of their time in individual or informal group study, and as the library does not provide some of the specialist physics study resources, a room particularly set up for students to study first year physics seemed desirable. A learning aids room was established for first year physics students. It contained three electronic calculators (to assist in experimental analyses); three super 8 film loop projectors and more than 100 film loops[3] (for remedial

---

[3] A few film loops have associated sound cassettes specifically developed for the first year courses, hence a cassette player and earphones were also available.

work and interest); reference books and a few current copies of popular physics journals; interest articles and "humorous physics" on the noticeboards (for interest and attitudes). Also worked solutions from tests (remedial feedback) and all official notices were posted in this room. Within the constraints of general university security the room was open whenever possible (8.30 to 6.30 Monday to Friday; 9.00 to 12.00 noon Saturday).

A sixth teaching and learning activity is student assessment and this will be considered in detail in Section 6.3.

In addition to Appendix E, which provides details of the relationship between each specific objective and the teaching programme, Table 6.1 summarizes the emphasis given to each objective in the four formal teaching methods. The table also shows how each course tended to emphasize different objectives.

6.3 Objectives and testing

Assessment of student achievement with respect to the course objectives has a variety of purposes. These can be categorized in terms of the feedback cycles in Figures 4.1 and 4.2.

A. Feedback to students; student assessment can;

(i) provide an incentive or motivation for study,
(ii) spell out the content and objectives of a course in operational terms and familiarize students with associated assessment techniques,
(iii) provide opportunities for learning,
(iv) diagnose learner difficulties and provide checks on progress with respect to certain criteria and/or with respect to peers,
(v) indicate readiness for further learning or need for revision,
(vi) guide decisions on future education or employment.

B. <u>Feedback to teachers</u>; student assessment can;

- (i) indicate student difficulties and priorities for future teaching,
- (ii) assess course effectiveness and hence stimulate changes in teaching and learning methods,
- (iii) lead to changes in the learning objectives,
- (iv) lead to modification of future student assessment.

C. In addition to these "intrinsic" purposes, student assessment can also be used to allocate grades for extrinsic purposes, and hence provide

<u>Feedback to the wider community</u>; student grades can;

- (i) indicate potential of students for further study or employment,
- (ii) act as a selection device for further study or employment,

and in certain cases

- (iii) encourage and support further study through high grades and scholarships.

It can be argued that student assessment during a course of study should be used solely to provide feedback to the student and the teacher. Terminal assessment at the end of the course can then emphasize the grading of students for the purposes outlined in C. In practice, student assessment during the course may also be used to assign grades. Reasons put forward for this are;

- (i) it enhances motivation and thereby possibly leads to increased attainment,
- (ii) it may increase the reliability and validity of the measure of student attainment,
- (iii) it enables students to gain immediate credit for work done and for "intermediate" attainments.

The use of student assessment during the course not only to provide feedback to students but also to assign grades for the purpose of indicating potential and for selection purposes can, however, introduce problems. There may well be considerable difference between

(a) what students know and what they can use at the end of a course (which is likely to be useful for feedback, prediction and selection), and

(b) how well they perform in specific tests and assignments during the course.

First, tests and assignments may emphasize particular objectives of the course, and ignore broad overall course objectives (for example, the synthesis of ideas from different content areas). Secondly, some students learn much from feedback from tests (or assignments) while others do not, due (for example) to differences in how much was correct initially or differences in student motivation. This subsequent learning is not assessed by the tests. Finally, first year students in particular often have different backgrounds and each student adjusts to the demands of a course in various ways with varying success. This adjustment involves changes in the motivation, knowledge, skills and attitudes of the students as the year progresses.

Another problem with assigning final grades in terms of internal assessment is related to student motivation. The type of internal assessment which may maximize student motivation towards achieving an objective of the course may not result in his submitting an assignment which can be used to assess his attainment with respect to that objective. For example, a set of problems may be given in physics as an assignment. In completing this assignment students undoubtedly learn some physics (first year Objective I) and hopefully increase their problem solving ability (first year Objective II). However, the assessment of this assignment is an assessment of the submitted result and may well be an indication of the amount of effort that the student has put into his work, rather than an indication of the extent of his working knowledge or his problem solving ability. Further, the student's knowledge, skills and attitudes may well be increased, perhaps even optimally, by open collaboration with other students, and this is desirable. However, such collaboration may well "inflate" his abilities as assessed by the assignment (essay or report).[4] Unfortunately it is tempting for the teacher to use internal assessment procedures which will increase student motivation rather than procedures which will result in the allocation of "appropriate" grades, even though the results will be used for the latter purpose.

From a slightly different point of view, the problem is that assessment procedures designed to measure the attainment of certain objectives (while providing useful feedback for teacher and student) do not necessarily provide maximum motivation for the achievement of these objectives. In the case of problem solving, on course tests which contain problems different from those previously prescribed as exercises should indicate a student's problem solving ability. However, students unable to solve these fresh problems after working hard at previously assigned exercises, are liable to quickly lose any motivation they have to develop their problem solving ability by working on subsequently assigned exercises. If test questions are set which are identical to those set for prescribed study, there is increased motivation for students to attempt subsequent prescribed problems; but in that case the grade attained from these test questions is likely to be measuring something other than problem solving ability. This problem does not apply to all objectives but probably only to those

---

[4] In addition if assignments rather than tests are used during the year a final examination provides a new evaluation procedure of which students have had, at least at that level, no prior experience.

objectives where something more than just industry is required to achieve them.

Obviously the advantages and disadvantages of using internal assessment not only to provide feedback but also to contribute toward the final grade will vary from course to course. In certain situations the most suitable type of assessment of the achievement of certain objectives is not appropriate for terminal assessment and any disadvantage of internal assessment will be offset by advantages in motivation, feedback and even validity. In the first year physics courses, because of

(i) the importance of providing guidance for future study,
(ii) the importance of the grades as a selection instrument,
(iii) the variation in student background, and
(iv) the nature of the objectives.

the weighting on internal assessment has been kept to a minimal 33% of the total course grade. It was decided that the terminal assessment of the achievement of the course objectives could be adequately covered by open book examinations. To ensure students had early experience of this type of assessment and to give them a clear indication of how objectives would be assessed in the final examination, all internal assessment which counted toward the final grade was in the form of tests of a standard and type similar to the final examinations. The two 2 hour tests in the experimental course, and the four 1 hour tests in the theoretical course, covered the type of questions to be found in the final examination and also covered the same objectives as that examination. Both final examinations were of three hours duration, each consisting of 30 multiple choice questions, 9 short (six minute) questions and 3 longer (twenty minute) questions.

In addition to the formal tests which contributed towards the final grade, other assessment procedures were included in the first year physics courses. These were used primarily for student self-assessment and they did not affect the student's final grade. In the theoretical course (73.101) the multiple choice tests in the evaluation sessions provided a means for students to check on their progress each week, while in the experimental course, self-evaluation questions were provided to enable students to gain some feedback on their competencies after each experiment (see Appendix D.4). As these results did not affect student grades, questions primarily designed to increase motivation, or boost confidence could be used. However, students can be critical of such questions if they are not similar in type and standard to those used in the more formal assessment situations and, in general, questions typical of formal assessment questions were used.

It is possible to use during-the-course assessment to manage the whole learning process, by making subsequent learning completely dependent on the results of such assessment. In the mastery model of learning and testing, each completed unit of work is followed by a "mastery" test. Progress to a subsequent unit of work depends on a student showing, via the test, that he has "mastered" a particular unit of work. Usually specific behavioural objectives are used so that a prescribed level of mastery can be specified for each objective (see Keller, 1968).

In the first year courses a mastery-managed approach to learning was not considered appropriate. Some of the problems associated with this approach include:

(i) If mastery testing manages the learning processes then the different times required by different students to reach mastery level (students resit tests until they pass before progressing) results in the necessity for individualized instruction. This "self-paced" instruction is not geared to a final grade and yearly progressions. Furthermore individualized instruction and mastery testing requires considerable resources. (A recent report from M.I.T. shows it may become quite impractical with large classes without great expertise (Friedman *et al.*, 1975).)

(ii) Techniques for developing criteria-based tests tend to be lacking and the concept of "mastery" is not particularly appropriate for a physics course with a variety of objectives. While for simple cognitive objectives it may be possible to discriminate between "knowing" and "not knowing" (mastery and non-mastery) there are no doubt various levels of understanding in the case of more complex concepts. Also skills and attitudes, such as the ability to solve problems, to think critically and creatively, and to find physics satisfying tend to develop gradually rather than be mastered in stages. In practice "mastery" can be a misnomer in that often mastery tests simply measure low level objectives, a "competency to proceed", or act as a "confidence booster". However in such cases, problems of the relationship between "mastery" tests and terminal assessment arise, as Higgenbottom (1975) recently pointed out. While criteria-based tests of low level objectives can be useful for feedback, peer related assessment over the broad range of objectives is necessary to assist students to make informed decisions regarding subsequent education and future careers.

In the first year physics courses assessment was neither purely criteria-related nor purely peer-related. Overall criteria standards tended to be based on the examiner's knowledge of the teaching situation and on his subjective expectations.[5] Scaling did occur where it was

---

[5] Subjective expectations generally developed from knowledge of student performances in tests and examinations in previous years.

considered desirable but not to a predetermined mean and standard deviation. In tests pass-fail boundaries were not specified (25% was better than 12%; 85% was better than 60%). For the final course grade students were graded A++, A+, A, A-, B+, B, B-, C+, C, C-, D+, D, E or X (absent). The pass-fail line was "obscured" by C pass, C- restricted pass and D+ opportunity for a further examination.

In terms of objectives and testing, Mueller (1973) has summarized possible variations in instructional models. An adapted form of this outline appears in Table 6.2. The first year courses, for the reasons outlined so far, can be described with respect to this table as primarily a "time unit fixed-all objectives to all students-normative referenced-normal curve assumed-maximum differentiation-general objective" type of instruction.

CHAPTER 7

THE EVALUATION OF INSTRUCTIONAL METHODS (PROCESS EVALUATION)

OUTLINE: It is argued that to evaluate instruction, particularly where the class is large, it is necessary to supplement informal feedback from students with more formal feedback procedures. Examples are provided of how in the first year courses, formal feedback procedures assisted decision-making about changes to instruction, and subsequently established the effect of these changes. Consideration is also given to the need for long-term monitoring of student reaction and to how more sophisticated evaluation instruments can be developed to evaluate particular aspects of a course.

7.1 Introduction

Once a teaching programme has been designed and implemented, evaluation is necessary to provide feedback for the subsequent improvement of the teaching programme. Process evaluation assesses the extent to which procedures are operating as intended. Product evaluation is concerned with the extent to which objectives have been, or are being, attained. While both types of evaluation provide feedback for controlling and evolving changes in the teaching programme, many teachers and curriculum developers tend to be more concerned with process evaluation than with product evaluation.[1] Two reasons for this are;

(i) process evaluation is closely related to the instructional procedures and can provide specific feedback for the making of decisions about instructional change and modification; and

---

[1] Particularly process evaluation in terms of student reaction to the learning procedures.

(ii) there are problems in measuring reliably the achievement of course goals in terms of absolute levels of student attainment or of changes in achievement, and in the relating of these measurements back to specific instructional procedures.

One of the strengths of a competent teacher is undoubtedly his ability to informally (often intuitively and almost subconsciously) collect, interpret and react correctly to day-to-day, even moment-to-moment feedback information from student reactions, comments, and actions. However, it is frequently useful to supplement this information with data acquired from questionnaires, rating scales, open-ended reaction sheets, semantic differential techniques, external evaluators and so on. This more formal type of process evaluation is of special importance where a large number of students are involved in a teaching programme.

In the first year physics courses, a number of substantive changes to the initial courses resulted from informal feedback. Tutorials, and particularly laboratory work, provided opportunities for staff-student exchange of ideas about the courses. In addition, more formal process evaluation procedures were developed.

A problem common to many aspects of process evaluation is that the act of measuring an attribute of a system may alter that attribute, or even a number of the system's attributes. This is particularly true if formal evaluation procedures make extensive demands on student time. With large student numbers sampling techniques can be used to reduce the demands on any one student. However, the number of first year students was insufficiently large to warrant this sampling technique. In the first year physics courses the total time spent by students during the year on questionnaires was approximately one hour.

The results of one isolated formal process evaluation are difficult to interpret unless responses are particularly positive (or negative), or can be compared with those relating to an alternative learning strategy. For this reason a continual monitoring of the educational system was considered desirable, so that significant changes could be observed and interpreted in terms of changes made to the teaching and learning activities.

So that responses to course surveys could be obtained from all the students actively involved in the courses, teaching time was used for the completion of questionnaires. For example, an experimental session would be interrupted, or an evaluation session curtailed, and survey sheets distributed with a brief comment and ten to fifteen minutes allowed for completion. While this technique did not artificially inflate positive responses, students were co-operative and it did ensure that responses were obtained from all present.

While forced responses (choosing between alternatives) were considered necessary to obtain a quantitative measure of overall student opinion, students were also encouraged to provide constructive free response comment. To aid the interpretation of a particular student response, the students were invited to identify themselves by signing their replies in those cases where this was not going to jeopardize honest replies (normally over 60% signed). Each year following the mid-year survey in the theoretical course, the percentage responses on the alternatives to each question were placed on the students' noticeboard. This enabled students to gauge overall student opinion and to follow this up, if desired, by discussion with other students and staff.

A typical survey sheet (end of year 1974) is provided in Appendix F.

7.2 Course process evaluation and major change (1970-1972)

In both the first year courses the textbook was envisaged as having a central rather than the more normal supportive, or supplementary, role. There was a limited range of textbooks in 1970 which contained the desirable content and philosophical orientation (stressing the unity of physics), which used the correct system of units (S.I.), and which was of a suitable mathematical level. Fundamental University Physics Vols I and II by Alonso and Finn (1967), was chosen as the best available text which met these criteria. However, it soon became apparent that this text was too difficult. Evidence came from informal discussions with students and from formal process evaluation (see Table 7.1).

During 1970, Physics by Alonso and Finn (1970) was published. This text retained all the features of the previous text but was considered to be of a more suitable standard. It was adopted in 1971. Formal process evaluation (see Table 7.1) showed that the percentage of students who considered the text "difficult" or "very difficult" dropped from 70% in 1970 to 23% in 1971. There was no significant change in the description of the textbook as primarily informative but not particularly interesting. These responses were considered acceptable.[2]

During 1970, students in the lecture-based theoretical course (73.101) were encouraged to concentrate on the lectures, not to take notes and to accept their textbook as fully detailed lecture notes (the reasons for this approach are provided in Appendix A.4). Although it was expected that the students might wish to annotate their textbook occasionally as the lecture proceeded, it was found that in practice many of them spent most of the lecture trying to ascertain the relevant place where the lecturer was in the textbook

---

[2] The data of Table 7.1 was supported by mid-year surveys which are not shown.

and underlining almost every word, equation or even symbol the lecturer referred to. In addition, some students annotated every minor change, and added alternative derivations, worked examples, and every extra detail mentioned. Any advantage to be gained from concentrating on the lectures rather than on taking notes seemed to be lost. Despite lectures based on their textbook, students in general did not have enough confidence in their own ability to cope adequately with the textbook as their only form of lecture notes. Even though specifically encouraged to do so, not more than one or two students out of ninety were able to sit back and concentrate on the lectures. This suggests that the students were conditioned to note-taking by early schooling and were unable to adapt to this method of learning.[3]

Formal process evaluation on the use of the textbook as lecture notes was obtained in 1970. At mid-year students were still reasonably convinced by the argument put forward for minimal note-taking, but they were somewhat less convinced by the end of the year (see Table 7.2(a)). In retrospect the response may have been more positive had the text been of a more suitable standard and had greater emphasis been given to the fact that the assessment questions would be developed from the textbook rather than from lecturers' notes. However, the decision was made following the 1970 experience not to discourage students from note-taking in 1971. Although students were aware that lectures were closely related to textbook, and that the tests and examinations in the theoretical course would be based on textbook only (with annotations), students invariably took notes in lectures in 1971. The 1971 percentage

---

[3] One successful student who sat in the front row and concentrated without note-taking was a repeat student from another university. He liked this learning situation. However, it is probably significant that the content was not new to him. The problem of note-taking was also discussed with a senior physics lecturer at another university. His reaction was probably not atypical--"but students learn through their fingers".

responses to the reciprocal question of Table 7.2(a) is provided in Table 7.2(b), column 1.

While restricted responses are desirable for comparison purposes, considerable caution must be placed on the interpretation of such responses in isolation because they are very dependent on the alternatives provided. In 1971 it was clear from additional free responses that there was a further reason why note-taking was desirable. This was included as alternative (E) in 1972 with a significant change in response from 1971.[4]

During 1970-1972 neither the formal nor the informal process evaluation suggested it was worthwhile to discourage students from note-taking. Research by Hartley and Marshall (1974) seems to support this view.

Students were asked their opinion about the novel procedure of using their textbook in tests and examinations (see Table 7.3). Similar distribution of responses were obtained in both June and October surveys in 1970, 1971 and 1972. Because of the reasons qualifying the "advantage" and "no advantage" alternatives, the results are of restricted generality. However, they do indicate an overall positive reaction to open textbook examinations and no change in the instructional system resulted from these findings.

The difficulty of interpreting formal feedback correctly, particularly if it is an isolated measurement, is exemplified in the student reaction to the question; how frequently should tutorials be held? In 1970 one tutorial was held per fortnight. Formal process evaluation resulted in the responses of Table 7.4. In 1971, despite formal feedback in 1970 suggesting that one tutorial per fortnight was

---

[4] No conscious change had been made from 1971 to 1972 in any related aspects of the instructional system.

considered by students to be appropriate, informal feedback resulted in a decision to increase tutorials to one per week. Formal student reaction to this (see Table 7.4) showed that, in fact, this was much more favourably received. The decision to change based on informal feedback, would seem to have been a desirable change. Further monitoring of the instructional system from 1972 to 1974 showed little change from 1971 in student percentage responses to this question and tutorials have remained at one per week.

Other aspects of the theoretical course were also changed in 1971 and 1972 as a result of informal and formal process evaluation. Student reactions to these innovations were obtained. For example, in 1970 the course was undoubtedly too difficult and the work load expected of students too heavy. In 1971 the content was reduced somewhat and the evaluation session (discussed in Section 6.2) was introduced to replace one of the three lectures each week. The effect of this and other minor changes such as providing summaries, multiple choice questions and discussion questions are discussed in detail in Appendix G.1. What is of significance is that the formal process evaluation indicated what changes should be made to the system, provided quantitative reaction of students to subsequent innovations, and provided quantitative evidence that desirable changes had resulted.

Process evaluation of the experimental course (73.102) in 1970 was limited, as the lack of experimental equipment necessitated an interim instructional programme for that year. Much of the 1971 and 1972 evaluation was concerned with the time required by students to complete each experiment and the problems they had with particular experiments. Because of the extensive resources required to change many experiments at the one time, formal process evaluation was primarily concerned in 1971 and 1972 with identifying the worst aspects of the learning system. The students were asked to list the experiment which they considered most difficult because;

(a) the theory behind the experiment was difficult to understand,

(b) the apparatus was considered inadequate,

(c) the time allowed for the activities was inadequate,

(d) the experiment proved difficult to write up adequately and satisfactorily.

In addition students were asked to assess the time spent in post laboratory study (ideally 3-6 hours), the interest level, and the information level of each experiment in the course. From this formal evaluation, as well as from information acquired informally, experiments were modified and changed. In Table 7.5 typical formal feedback information for one experiment is provided for 1971 and 1972 as an example. This experiment was on transient currents and a.c. theory and the 1971 data suggested that the experiment was too long, of high information, and of average interest. Of the twelve 1971 experiments, approximately 25% of students considered this experiment the most difficult in terms of (a) and (d) above. Consequently, the experiment was modified for 1972 and feedback then showed that only 10% of students considered it the most difficult in terms of (a) and (d). Further, student responses shown in Table 7.5 indicated that when the length of time spent on post-laboratory study was reduced, the interest level was increased. However, these changes were made at the expense of the information level. This problem of competing objectives (in this case between cognitive and attitudinal objectives) in course development was found to be not uncommon and will be considered further later.

A number of other aspects of the experimental course were investigated in 1971 and 1972. These investigations showed, for example, that the detail provided in the experimental guide was at the desired level; that student reaction to a six-hour laboratory per fortnight rather than a three-hour laboratory per week was favourable; how students were using the self-evaluation questions provided; and what weighting students considered should be given to their laboratory notebook in student assessment. Details of these are provided in Appendix G.2 where, in addition, the problem of persuading students to write directly into their notebooks is discussed.

In the experimental course, as in the theoretical course, free student responses were encouraged. Most comments were related to process evaluation and invariably were sensible mature comments. Such comments can be useful for course improvement. However, they are difficult to classify and it is also difficult to assess the likely sympathy of all students with a particular individual's comment. Free responses can be particularly useful for the purpose of designing subsequent surveys, choosing pertinent questions, and deciding on response alternatives to be included in subsequent fixed response surveys. However, with on-going monitoring, the desirability of improving a survey has to be considered against the difficulty of comparing student responses from one year to the next if surveys are different.

7.3 Process evaluation and course monitoring

Formal process evaluation enables aspects of a course to be continually monitored. In this way gradual but significant changes in the instructional system can be detected and acted upon. In 1970 the work load in the theoretical course (73.101) was excessive and it was significantly reduced for 1971 (see Appendix G, Table G.1). Figure 7.1 summarizes the variation in student response to the theoretical course work load from 1971 to 1974. A significant reduction in work load was again made from 1971 to 1972. Subsequent changes are of little statistical significance although the gradual increase of the perceived work load from October 1972 to October 1974 indicates the desirability of continual monitoring. The absolute level of student response is not useful information unless it can be compared with identical investigations in other subjects. However, equivalent data for the experimental course (73.102) for October 1974 shown in Figure 7.1 indicates a need for a further reduction of the work load in that course.

A second aspect of a course which needs to be monitored is the content and pace of lectures. This need is clear in cases where the background of students varies or where different lecturers are presenting the lectures each year. However, even if the lectures are presented by the same lecturer each year it is possible for unconscious changes in the lecturing to occur even apart from deliberate changes in content. Student reaction to the content and pace of lectures over the five-year period from 1970 to 1974 is summarized in Figure 7.2. As anticipated, because of considerable changes made to the theoretical course, there was a consequential change in apparent difficulty from 1970 to 1971. From 1971 to 1973 no deliberate changes of any moment were made in the lecturing style. However, during this time both the difficulty and pace of lectures increased. The background attainment of the student responders improved during this period,[5] and the apparent difficulty of the text remained static. These two facts suggest that the changes were lecturer rather than student dependent. Possible reasons for this change could be;

(i) with "better" appreciation of student difficulties there was a tendency to emphasize difficult content rather than spend time on introductory concepts which now seemed

---

[5] The percentage of students who had a seventh form physics background rose steadily from 68% to 80% from 1970 to 1974.

straightforward to the lecturer. Possibly III(B) and III(C) objectives came to be emphasized rather than I(B) and I(C) objectives,

(ii) there was a tendency not to "turn the clock back" sufficiently from one year to the next, i.e., not to appreciate that the entrant students each year were not comparable to the first year students taught just prior to the end of the year.

Whatever the reason a particular effort was made in 1974 to reduce both the difficulty and pace of some of the lectures. The perceived relative difficulty of the chapters in the text had been monitored from 1970 to 1972, and in 1973 information on how difficult students found each lecturer's section of the content was collected. As seen in Figure 7.2 and Figure 7.3, student responses in 1974 improved with respect to perceived course difficulty although students' perception as to the pace of lectures remained high.

Another problem identified by process evaluation was the frustration students experienced in attempting to solve problems. Despite the fact that students found the text to be of a reasonable standard, and that much guidance was given to students in evaluation sessions and tutorials, they still found suggested problems "impossible" (5%) or at least "difficult" (55%). Although problems selected for students were reconsidered each year, still 50% of students in 1974 considered problems to be "difficult" rather than "of a suitable standard". According to the June surveys typically only 50% of students attempted most problems set, and this percentage reduced to nearer 40% by October in most years.

Consistently, investigations into why students had "trouble" with problems showed (1970-1974) that 70% of students "could not see how to start the problem". The whole problem was typified by an overseas student "I do not attempt problems now because in my case anyway I get all set to do some physics problems and get stopped because I cannot get started."

While there is a full awareness of the problem, no solution has yet been found which significantly affects student responses. Tutorials were used extensively to assist students with problems and typically 70% of students stated tutorials were usually "helpful" versus "useful only if you have questions to ask."[6] A booklet was produced by a staff member in 1974 (see Pepper, 1974) to help students develop successful procedures for solving problems and to boost confidence. In retrospect simpler problems in the course and a greater inducement to attempt these problems may have been appropriate for many students. However, there is a real conflict in that greater inducement is likely to increase the work load to an unacceptable level.

As an example of an otherwise undetected trend, Figure 7.4 shows that student attitude to the interest value of the textbook Physics by Alonso and Finn (1970) has declined significantly since 1971 when it was first introduced. In contrast the perceived difficulty of the text and attitudes to some other aspects of the course have not shown this decline. Possible reasons for the decline in textbook popularity could be the possible use by secondary school teachers of the textbook in their pre-university teaching, the wider societal changes in attitudes towards physical science in recent years, or the fact that the book is no longer a recently published text and modern texts in other subjects are becoming more attractive. Whatever the reason this is an example of a trend which needs further monitoring and the continual appraisal of new texts as possible alternatives.

---

[6] In the surveys 1971 to 1974 typically only 2% of students chose the third alternative, "a waste of time".

Continual monitoring of aspects of the experimental course were not undertaken in the same way as in the theoretical course. As mentioned previously, the problem of obtaining and developing suitable equipment not only meant that the course took longer to develop initially, but it also meant that only limited changes could be subsequently made without making excessive demands on manpower or finance. A policy of changing significantly not more than one experiment per year (≈ 10% of the course) was adopted in 1972. Consequently the prime concern of process evaluation was to identify each year the least successful experiments. Even this is difficult. In terms of student attitude alone a "successful experiment" is not a unidimensional quantity. A Spearman's rank order correlation of - 0.1 was obtained in 1972 between students' rank order of experiments on an "interest" scale and on an "information" scale. As one student stated, "I enjoy an experiment for which I understand the theory. However I learn little new. An experiment requiring me to understand theory I have not met before results in an informative experiment, but usually I do not consider it particularly interesting."

As far as it was possible, without placing undue demands on students, an attempt was made to gain a better understanding of the attitude of students to experiments and to develop an appropriate instrument by which these attitudes could be measured. This will be discussed in Section 7.5.

Comment has already been made on the importance of informal discussion with students in terms of process evaluation. Similar informal evaluation was carried out by the University of Waikato's 1973 Secondary School Teachers' Fellow in Physics. This experienced teacher attended all the first year physics lectures, and was involved in tutorials and laboratory work. The day-to-day informal comment of such an "external evaluator" cannot be overestimated.

7.4 Process evaluation and the objectives

Much of the process evaluation considered in Section 7.2 and Section 7.3 was not explicitly concerned with course objectives. Rather the evaluation was of specific learning activities, e.g., lectures, tutorials, and experiments. However, as mentioned in Section 6.2, these different learning activities focus on different objectives, and in this way the evaluation can be related to the objectives of the course. Problems with the difficulty level and pace of lectures are related to Objectives I, III(B) and III(C). Difficulties that students have with problem solving are related to Objective II. Students' attitude to discussion questions is connected with Objectives III(A) and VI. Summaries are primarily concerned with the achievement of Objectives I and III(C) and the fact that students found them helpful is useful process evaluation.

Students' understanding of the learning objectives is often not particularly well developed even by the end of the course, and survey questions referring to the objectives even when simple and direct are easily misinterpreted by students. Purely for interest however, and without any intention to base changes on the information provided, one set of questions was explicitly related to the objectives of the course. Students were asked which of the variety of learning environments in the courses was the most effective with respect to the objectives of the courses (see Appendix F, question 2). Students choosing more than one environment were given a proportionately decreased weighting on that objective, so that a 50% student response on learning environment A for objective X could be that 50% of students chose environment A only, or that all students chose environment A and F as both equally effective, and so on.

Similar data were obtained for 1973 and 1974 and Figure 7.5 shows the total relative weightings of the learning activities for each objective. The students' evaluation of the effectiveness of these learning environments can be compared with the actual percentage of time allocated to each activity. If it is assumed that a student attempting both courses should spend a total of 12 hours per week studying physics the division of his time would be: lectures 17%, tutorials 8%, laboratory studies 25%, evaluation sessions 8%, and informal study with friends and self-study 42%.

While this particular survey was not without ambiguity, student responses were not incompatible with the intended relationships between teaching and the objectives. For BASIC UNDERSTANDING students rated lectures higher than expected and laboratory studies somewhat lower, in spite of the emphasis on understanding in the experimental course. The relationship of the tutorials to the objectives was as anticipated, while the effectiveness of the evaluation session was lower than expected. The perceived effectiveness of laboratory studies for all but the first two objectives is significant.

Students were asked in October 1971 whether they considered they had "learnt more physics" from the theoretical or from the experimental course. Ten percent of students refused to choose either alternative stating that they considered the courses were equally effective. Typical responses were, "the courses tend to complement each other", "they are totally different types of learning". Thirty percent of students considered they had "learnt more physics" from the experimental course. Free response statements made by some of these students were, "experimental gives a deeper insight into certain phenomena than is generally gained in the theoretical treatment", "experimental helps you understand the application of theory", and "possibly because I enjoyed the experimental course more than the theoretical course, I got more out of it". The remaining 60% of students considered they "learnt more physics" in the theoretical course but less frequently supported their decision with a free response. However typical responses were, "the theoretical course starts at the beginning assuming you know very little or nothing, for experimental you need a lot of basic knowledge", "the theoretical course covers a wide range of material". It was clear that the decision of many students on this question depended on whether "learnt more" was interpreted in terms of depth (biased to experimental course) or breadth (biased to theoretical course).

In general student responses did indicate that students appreciated the theoretical component of the experimental course and did not view the experimental course simply as a practical course.

7.5 Developing a process evaluation instrument

As stated earlier (see Section 7.3), an on-going policy of minor (incremental) change was proposed for the experimental course. One of the major requirements of process evaluation in this situation is the identification of unsuccessful experiments requiring modification or complete replacement. Many aspects of experiments can be evaluated, e.g., difficulty, length, information level, satisfaction level, importance in terms of content covered, importance in terms of all or some of the course objectives, and so on.

Many students indicated that the laboratory studies were an effective learning environment for developing an interest in physics and for appreciating the nature of the subject (Figure 7.5). Further a 1972 survey showed that approximately 50% of the students considered that the courses had improved their attitude toward physics as they felt more positive towards the subject, while less than 20% considered their attitude toward physics was more negative. This reinforced the view that incremental change was an appropriate policy, at least in terms of the attitudinal goals. An early investigation to establish the reasons for the students' change of attitudes showed that laboratory studies had a major influence on attitude change for both the "positive change" and the "negative change" groups of students (see Figure 7.6).

The above information suggested that for the first year physics courses, laboratory studies have an important effect on attitudinal goals and even on a student's appreciation of what physics is. As mentioned in Section 7.3, an analysis showed that there was no significant correlation between the students' perception of their interest in an experiment and their perception of its informational value. All this suggested that a more comprehensive study of student attitudes to experimental work was desirable. It was felt that such a study could lead to a determination of what aspects of an experiment are highly correlated and hence might establish what aspects are particularly associated with the enjoyment of the experiments, as well as with their informational value. The study might also result in the development of a more useful instrument for measuring student attitudes towards an experiment and for measuring changes in these attitudes.

The semantic differential technique developed by Osgood (1957) has a variety of uses in the field of attitude measurement and was considered an appropriate technique for this study. In 1973, and again in 1974, twenty-six pairs of bipolar adjectives were used to obtain the factor structure of student attitudes towards a set of experiments. In 1973 experiments 1, 2, 3A, 3B and 4; and in 1974 experiments 1, 2 and 3A (see Osborne, 1974a) provided the concepts to be considered and the twenty-six bipolar adjectives shown in Table 7.6 were chosen to cover attitudinal, informational and other parameters of the experiments.

The order of the adjectives was altered from concept to concept (experiment to experiment) and a sample of 53 students in 1973, and 80 students in 1974 who were currently engaged in the course completed the 26 seven step scales for each experiment. (A sample of the sheets given to the students is provided in Appendix H.) As the field of interest was the factor structure over a variety of experiments, the 26 x 26 correlation matrix summed over both students and concepts was obtained (N = 265 in 1973, 240 in 1974).

Each correlation matrix was analyzed using a principal components analysis (with unity in the diagonals) on both an IBM 1130 and a B6700.[7] The analyses yielded four significant factors and these were orthogonally rotated using Kaiser's Varimax rotation (Kaiser, 1968). The results are given in Table 7.7 where all loadings $\geqslant 0.5$ are shown and the bipolar adjectives have been ordered so that adjectives with comparable loadings are grouped together. The proportion of the total variance accounted for by the first four factors was 0.56 in 1973 (0.54 in 1974).

As Kane (1969) has re-emphasized, the meanings of scale defining adjectives may change from concept to concept and the factorial structure of Table 7.7 would not necessarily be invariant over other sets of experiments, let alone other educational programmes. However, for the particular decision setting for which such an instrument might be developed and used, the generality of the factor structure would not be important. Nevertheless, the similarity of factor structure over a slightly different set of experiments with different students indicates the reliability of the analysis and affirms that the factorial structure is reasonably stable.

Adjectives loading heavily on Factor I suggest that this is an enjoyment factor and it is interesting to note that 'varied-monotonous' is loaded on this factor. Factor II might be described as a value factor. The particular bipolar adjectives which have heavy loadings

---

[7] In 1973 an orthogonal component factor analysis (with the square of the multiple correlation in the diagonals) was also computed. Loadings were similar to the principal components analysis.

on Factor III suggest that this factor is related to the difficulty of an experiment while Factor IV appears to be associated with the mathematical complexity of an experiment.

To provide comparative information about experiments, a profile of the mean student values for each of the 26 bipolar adjectives was obtained, as shown in Figure 7.7. From a comparison of the profiles of different experiments significantly low values of particular means can be identified. A knowledge of the factor structure for the adjectives gives some indication of the likely effect on attitudes of changing one aspect of an experiment. For example, it seems likely that a greater variety of activities in an experiment will increase the enjoyment of that experiment. Figure 7.7 does show that significant differences do occur between experiments, so that a new experiment could be expected to show a significantly different profile from the replaced one. The stability of the profile from one year to the next, if an experiment is not changed, is shown in Figure 7.8 (where an overall more positive attitude of + 0.3 by the 1974 students over all experiments has been allowed for). This stability in profile shape gives an indication of the reliability of the profile. While no changes were made to Experiment 2 from 1973 to 1974, some modifications were made to Experiment 1 at the end of 1973.[8] These instructional changes appear to have had little major effect on the profile (see Figure 7.9). However, significant positive changes ($p \leqslant 0.01$) in the absolute mean values from 1973 to 1974 did occur for three of the Experiment 1 bipolar adjectives; these were short-long; straightforward

---

[8] The second half of Experiment 1 was restructured to give greater emphasis to the relationship between experiment and theory. Rather than use a spherometer to measure the radius of curvature of a spherical dish the method was left more open and students were encouraged to use geometric optics to determine the radius. It was hoped that this might add variety to the experiment, reduce the unnecessary complexity introduced by using the spherometer and revise some geometric optics.

theory-complex theory; varied-monotonous. No such significant change occurred in any of the mean values associated with Experiment 2. The nature of the "significant" adjectives is not inconsistent with changes made to the experiments, and the results do suggest that a semantic differential survey of the type detailed could provide a valid measure of attitudinal change to experiments.

One of the major problems with developing and using such an instrument is that if a large number of experiments is to be covered the number of bipolar adjectives would need to be considerably reduced. It is not reasonable to expect students to complete more than a total of about 100 bipolar adjectives in one sitting. However, once the factor structure has been determined for a set of experiments, a limited number of adjectives representative of particular attitudinal dimensions could be used to provide a much more limited number of bipolar adjectives. Finally, it is not envisaged that such an instrument would provide more than supplementary evidence, which would in conjunction with all the other evidence available, help in the making of decisions about what changes to experiments are desirable.

In conclusion, the last statement can be generalized with regard to the types of formal evaluation and monitoring procedures considered in this chapter on process evaluation. The procedures basically provide valuable supplementary evidence to assist in deciding what modifications, or changes, are to be made to the instruction.

CHAPTER 8

THE MEASUREMENT OF STUDENT ACHIEVEMENT (PRODUCT EVALUATION)

OUTLINE: The complex relationships between product evaluation, objectives, and student assessment, which exists where the objectives of a course are expressed in terms of student behaviour, are explored. Consideration is given to how best to measure student achievement of individual course objectives. Details are provided of the development of achievement measures and of subsequent statistical analyses of measured student achievement.

8.1 Introduction

Unlike process evaluation which in this study has been concerned with investigating directly the instructional methods and how they might be "improved", product evaluation is more concerned with the extent to which objectives have been, or are being, attained. Product evaluation is simplest when the objectives of the course are specific. For example, if an objective of a course is to reduce the dropout rate from 40% to less than 10%, a clear product evaluation measure is identifiable, and the results obtained are unambiguous. If an objective of a course is that students be able to identify 10 named parts of an electric motor, again product evaluation can give clear specific answers as to whether the objective has been achieved or not. If, however, the objectives of the course are more general (e.g., to develop a student's ability to solve problems, to think critically, and so on) then although these objectives may be considered more appropriate goals for learning, product evaluation will not yield clear-cut answers.

As stated in Section 5.2, it is desirable to express the course objectives in terms of student behaviour in that this clarifies the purpose of instruction, and guides teaching and student assessment.

Hence in terms of the above, very specific behavioural objectives are desirable for product evaluation as criteria-referenced measures can be devised which are related to these objectives. However, although product evaluation becomes simpler as objectives become more trivial, context evaluation becomes more difficult as the question, "what is the value of the objective?" (for which the achievements can be demonstrated) becomes more pertinent. Criteria-referenced tests may tell what an examinee knows or does not know; however, what they do not tell is how good or how poor his knowledge or ability may be (Ebel, 1971). Excellence and deficiency are necessarily relative concepts.

In addition to the apparent desirability of specific behavioural objectives and criteria-referenced tests for product evaluation, it may seem desirable to use pre-test and post-test to measure student's gain-score resulting from instruction, particularly to evaluate course effectiveness. However, there are certain doubtful assumptions behind such a procedure as Roebuck (1972) points out. The first assumption is that the tests which measure the pre- and post-test proficiency can be equivalent—however, even an identical test no longer measures the same thing after instruction as it did before instruction. Secondly, there is the assumption that the scale units are uniform throughout the range of possible gain-scores for a given test—however, a gain from 40% to 50%, for example, is not equal to a gain from 70% to 80% for a variety of reasons including the ceiling effect. Thirdly, the assumption is that gain-scores are valid and reliable. However, the unreliability of pre- and post-tests is compounded in gain scores and there exists a spurious negative correlation between pre-test scores and gain-scores. Falk and Dow (1971) also point out that a student's gain-score can only be used with caution as evidence that the course produced the learning. For example, teaching in one year can influence learning in the subsequent year in a way not assessed by the pre-test;

competent students are likely to learn in spite of bad teaching.

In practice, when the objectives of a course are written in terms of student achievement, whatever the level of specificity, product evaluation becomes associated with student assessment. However, assessment procedures which measure student achievement serve a wide variety of purposes in addition to measuring course effectivess (as listed in Section 6.3). In particular, end-of-course tests, examinations or other procedures whether at the national or classroom level normally must be concerned primarily with the assessment of individual student behaviour and the differentiation of students for selection purposes. Whether such assessment can also be used to evaluate the effectiveness of a course in attaining its objectives is open to some debate. Undoubtedly it depends to some extent on the nature of the objectives of the course. Kelly (1971a), in the trials for Nuffield A-level biology in Britain, assumed that there need be no conflict between developing an examination to measure attainment in a way suitable for selecting students for higher education and developing an examination to evaluate the effectivenss of a course. On the other hand, Roebuck (1972) emphasizes that achievement tests are used for different purposes and suggests that particular uses may require tests with particular sensitivities.

It seems likely that in a teacher-operated system where the objectives of the course are written in terms of student achievement, product evaluation has to be viewed as largely a by-product of "official" student assessment. In certain situations it may be desirable to use supplementary testing procedures particularly designed for product evaluation. However, if the course objectives are closely associated with final student achievement it is difficult to justify such a procedure.[1]

Student assessment closely related to course objectives, irrespective of how specific these objectives may be, can provide valuable feedback information for product evaluation. In the formative stages of course development, as test questions are designed to measure, or at least indicate,[2] student achievement of certain objectives, the objectives are further clarified both to the teacher and to the student. At the same time the examiner is extending his experience and his knowledge about how to assess the achievement of the course objectives. As difficulty-levels are adjusted to provide suitable questions to assess student achievements (normative-referencing), some consideration, albeit subjective, can be given to the actual standard being attained (criteria-referencing). If the standard is considered unsatisfactory (either on internal or external grounds) steps can be taken to review the learning processes and/or the students' efforts.

To some extent the necessity for the scaling of normative-referenced examination papers either internally (by the examiner) or externally (by an examinations board) gives an indication of any significant differences between the examiner's expectations and students' attainment. However, the necessity for scaling could just as likely be due to unreasonable expectations on the part of the examiner, as to course ineffectiveness. Conversely, however, if an examiner considers his paper reasonable (and he is supported in this view by colleagues familiar with the course or by external moderation) and

---

[1] Apart from measuring attitudinal objectives which students would not want to be part of the official assessment.

[2] Lord and Novick (1968) distinguish between measures and indicators of theoretical constructs.

little scaling is required, then at least at a subjective level the course is apparently achieving the measured goals at a reasonable level (assuming marking is rigorous).

With regard to product evaluation of the first year physics courses, it was decided that the examinations and tests would attempt to measure all the non-attitudinal objectives of the courses (i.e., excluding V(C) and VI objectives). However, in the case of any conflict of purpose, it was agreed that attainment would be measured primarily in a way suitable for selecting students for advancing toward their degree. It was decided that it was not desirable or practical to use identical pre-tests and post-tests to measure course effectiveness, nor to concentrate on purely criteria-based measures,[3] nor to develop alternative product evaluation measures except in those areas where the objectives could not be assessed by selection procedures (e.g., attitudes). While this approach to product evaluation of course effectiveness has inadequacies, no feasible absolute measures are available at present. The above approach, however, does provide opportunities to investigate the problems involved in measuring student achievement of course objectives; to relate objectives to teaching and testing; and to obtain useful information for course evaluation and improvement, and data for model building. Further it places emphasis on student assessment which is an inherent part of any instructional system, and which must be understood thoroughly before really adequate product evaluation of courses will be possible.

8.2 The development of achievement measures

With respect to the first year physics courses, tests and final

---

[3] As Roebuck (1972) pointed out the techniques for the production of such tests and the efficacy of them in practice have been relatively unexplained.

examination papers were usually developed without any pre-testing of the questions, primarily because of security problems. However, every paper developed by a staff member was closely scrutinized by at least one other staff member in terms of face validity and likely difficulty level (including the suitability of the time allocation). During 1970-1972 all multiple-choice tests were subsequently statistically analyzed, and difficulty levels of, and correlations between, free format answers were also investigated. Useful information was obtained from these analyses and experience was gained in predicting the likely difficulty of proposed questions testing particular objectives and the likely discrimination of particular questions with respect to the overall test mark. A study of individual answers, and a study of the way distractors in multiple-choice tests operated, helped to clarify questions and indicate whether or not questions were discriminating between students on intended objectives. In this way expertise was obtained in developing questions related to particular objectives.

As it was intended to develop a statistical model of the first year courses in 1973 (later extended to include 1974), particular attention was given during 1970-1971 to how best to measure achievement of individual course objectives. Correlation matrices derived from question-scores obtained in tests and examinations in 1971 and 1972 showed that a student score on an individual question is a complex function of content covered by the question, the position of the question in the paper, question type and objective being assessed. If student achievement of any one objective was to be investigated then a decision had to be made about how best to measure this achievement.

In teaching, decisions have to be made about what teaching methods are likely to be most suitable to develop the achievement of particular objectives. In the same way, decisions have to be made about what assessment procedures are likely to be most suitable to measure the achievement of particular objectives. For the development of a statistical model of the first year physics courses it was decided that as far as possible multiple-choice questions would be used to measure achievement.

Particular advantages of multiple-choice questions are that:

(i) Students can be more certain of what is expected of them. It is possible, therefore, to confine a question to a particular objective and avoid contamination of the question by requiring the use of irrelevant abilities. There is little doubt that the objective-type question leads to a focus on educational objectives (Cox, 1972).

(ii) Four or five multiple-choice questions extracted from a range of content areas enables a relatively content-free measure of the achievement of an objective. One short "free format" question (normally from a single content area), which may take students equally long to answer, cannot do this.

(iii) Questions can be pre-tested more adequately than other types of questions; a knowledge of how distractors are acting can provide clues as to whether or not the question is likely to assess the intended objective.

(iv) Problems of marker reliability can be reduced to a negligible level using cross-checking procedures.

Questions with free format answers (essay questions, short-answer questions) can only be judged suitable for the assessment of a particular objective if full details of the marking scheme are known. Information obtained from pre-testing such "subjective" questions is difficult to analyze and the assessment of unexpected answers to a question introduces other problems. However, for some objectives, particularly those requiring students to propose ideas, to create designs and to draw conclusions (divergent thinking—Guilford, 1971) multiple-choice tests were not considered appropriate and short questions requiring students to formulate their own answers (free format) were used.[4]

To develop measures of student achievement which could be used to construct a model of the instructional system in 1973, two test batteries were developed during 1971 and 1972 to measure the achievement of each course objective both prior to, and following, the first year courses. Questions were different for the two batteries and each question was specifically designed to assess an objective of the course in terms of the prior learning environment in which the student had been involved. For the input tests this was assumed to be seventh form physics and the "model" was developed solely for these students. The "input tests" were designed to be taken by students in the first week of term; the "output tests" were designed to be incorporated in the final examination papers.

Normally, to develop achievement tests, a large number of questions (items) are written and pre-tested on a group of students similar to those for whom the tests are being designed. Subsequently a set of questions is selected from this pool of pre-tested items on the basis of content area, the objective covered by the question, the difficulty level and the discrimination index of the question. The discrimination index approximates the value of the correlation of the question-score and the total-score on a "homogenous" set of pre-tested items. Ideally, all questions in the "homogenous" set should be on the same content area and have the same objective as the question being

---

[4] Normally where achievement of a student with respect to a particular objective was not to be investigated (i.e., in tests and in the 1970-1972 examinations) attainment was normally assessed both by free format and by multiple-choice questions.

considered. In practice, however, this is not always possible unless there are only a few objectives being pre-tested and there is a large number of pre-test questions. If the correlation is determined between the score on the question and a total mark obtained from a heterogenous set of pre-test questions covering a range of content and objectives, then the discrimination index must be viewed with caution. This is particularly so if it is suspected that student achievement of the different objectives are not necessarily highly correlated. Hence a low discrimination index does not necessarily mean the test item is invalid although other possible reasons for the low value should be checked, e.g., there may be a mistake in the question, students may have been incorrectly taught, or the question may be unduly easy or difficult.

In developing the "input" and "output" test batteries, each question was specifically written to discriminate as far as possible on one particular objective and the questions were categorized accordingly.[5] The objectives of Table 5.1 were assumed to be somewhat hierarchical within a particular concept. For example, Objective I was assumed to be a more "basic" objective than Objective III, Objective II(A) was assumed to be simpler than Objective II(B). However, in the case of any one physics concept, questions could be hierarchical simply because of the interdependence of the questions.[6]

Just as different learning situations provided different

---

[5] This categorization was reassessed in the light of pre-test data and sometimes this alone resulted in a question being excluded from the final test but occasionally the question was recategorized.

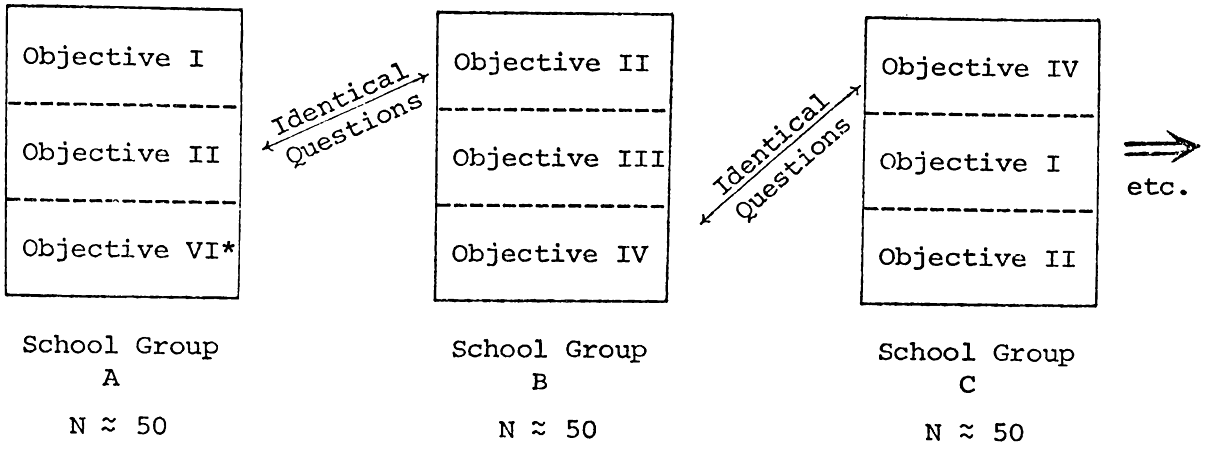
[6] An example of this hierarchy is seen in three multiple-choice questions used in a national examination by the author. The questions (Appendix I.1, Section II, Q.13, 14, 15) were designed to test Objectives I(B), II(B) and III(B) with respect to one particular situation. The facility levels of the questions were determined to be 76%, 40%, 24% and 85% of the scripts scored in the accepted hierarchical pattern (N = 500).

opportunities to achieve particular objectives, so different learning situations provided different opportunities for assessing the achievement of different objectives. Where content was new or difficult, questions attempting to measure Objectives I(B) or I(C) were most applicable and discriminated well between students. Where the basic content (Objective I) was familiar, opportunities to measure "high level abilities" existed. Each question was therefore categorized in terms of the most likely reason that students would be unable to answer the question correctly. For example, an Objective II(C) question was expected to discriminate not because of a lack of knowledge but because of the fact that the student did, or did not, realise that the knowledge available to him was relevant, or could be applied, to this problem situation.[7] Because the final assessment sampled a wide range of content, the fact that a student was unable to answer an Objective I-type question in one content area did not necessarily mean that he would be unable to answer an Objective III-type question in another content area. Hence any hierarchical order of objectives was submerged and not investigated by these tests.

To develop the "input" tests, 240 items were written and pre-tested in 23 seventh form classes (N = 300) in a variety of secondary schools in November, 1971. In this pre-testing phase, questions were grouped according to the objectives being tested. Different groups of questions were combined to form six different "tests" each consisting of about 60 cognitive items and 40 Likert scale items to measure attitudinal objectives (see Section 8.3). The objectives in the tests were lapped, as indicated below, so that each set of items on a particular objective appeared in two or three tests.

---

[7] The open book examinations decreased the likelihood of a memory lapse contaminating the measurement of the objective in question.



*Likert scales analyzed separately.

Data obtained from pre-testing the multiple-choice questions were analyzed using a program (W Mark) at the University of Waikato computer centre (IBM 1130).[8] Each "test" was analyzed as a whole, and also the results of performance on the same questions in one particular objective area were combined from all relevant "tests" and analyzed as a single group. Hence rather than obtain a single analysis of a large group of student responses, a number of analyses using smaller numbers of students provided multiple information which was tabulated for future reference. (See Appendix J for a typical analysis.) Similar analyses were made of the free format answers. The product moment correlation between scores on an individual question and the total score for a set of similar questions was obtained as well as facility indices. Analysis

---

[8] The W Mark program provides the mean, standard deviation and histogram of total scores for a specified group of questions. In each analysis, the "top" and "bottom" 27% of students chosen on their total mark for all questions is used to determine the facility index (F) and the discrimination index (D) for each question. If U and L are the numbers of students who chose the correct answer for a particular question in the top and bottom groups respectively, then for that question

$$F = \frac{U + L}{2N} \quad \text{and} \quad D = \frac{U - L}{N} \quad \text{where } N = 27\% \text{ of all students.}$$

In addition for each question the analysis provides the number of students choosing each distractor. This is obtained not only for all students, but also for the top and bottom quartiles, where the quartiles are chosen on the total mark on all questions.

of the Likert scales (used to measure attitudes) will be discussed in Section 8.3.

From the pre-test data, questions were selected for a battery of input tests. Each question was selected on face validity, content coverage, difficulty, and either a discrimination based on, or a correlation with, other pre-tested items intended to measure the same objective. Because each objective was not necessarily unidimensional not too much emphasis was given to the discrimination index unless it was low or negative. For most objectives the different discrimination indices, obtained from "across-objectives" analyses compared with "within-objective" analyses, were not widely disparate. However, where all the other reasons for selecting one question compared with another were equal, greater weighting was given to the "within-objective" discrimination index than to the "across-objectives" discrimination index.

The battery of input achievement measures related to the course objectives consisted of four M.C. questions for each objective of Table 5.1 except that;

(i) there were no questions on Objectives I(A), IV(D), and VII(A) as these were not considered applicable to the seventh form course,

(ii) Objectives IV(A), IV(B), and IV(F) were assessed by non-objective questions as it was assumed the achievement of these questions contained elements of divergent thinking,

(iii) Objectives V(C), VI(A), VI(B) and VI(C) were each assessed by Likert items (see Section 8.3) as this seemed more appropriate for these "attitudinal" objectives.

Approximately 40% of the 240 pre-tested items and questions were used in the test battery. The finalized input tests and analyses of individual items is provided in Appendix K.

Questions for the "output" assessment of each course objective (which will be called the "output questions") were chosen and modified from questions "pre-tested" in the 60 hours of normal tests and examinations given to first year physics students taking the courses in 1970, 1971 and 1972. Questions were selected in terms of content coverage, objectives and statistical information consistent with the objectives to be assessed.[9] For the development of a statistical model of the system (PART C), questions were chosen as similar in type as possible to the "input" questions for each objective.

Each objective was assessed by five multiple-choice questions except that;

(i) Objectives IV(A), IV(B), IV(C), IV(D) and IV(F) were each assessed by two 6 minute questions,

(ii) Objectives V(A) and V(B) were each assessed by two multiple-choice questions and one 6 minute question,

(iii) Objectives V(C), VI(A), VI(B) and VI(C) were each assessed by 8 Likert items used in the input tests.

The final examinations in the two first year physics courses each consisted of three 20 minute questions, nine 6 minute questions and thirty multiple-choice questions in 1973 and 1974 with no choice. The "output questions" (apart from those in (iii) above) were included within this format with the addition of some supplementary questions particularly in the theoretical examination. These supplementary questions were 6 minute and 20 minute questions which primarily emphasized Objectives I, II and III, but were neither pre-tested nor as precisely related to a particular objective as the "output questions".

---

[9] The experience gained from pre-testing "input test" items was invaluable in developing output questions where the pre-testing was limited because of security and the number of students in the courses.

Some supplementary questions changed from 1973 to 1974 although they covered the same content and the same objectives in both years.[10] In the development of a model of the first year physics courses (Chapters 10-13), these supplementary questions were not used for the purpose of measuring a student's achievement in any one objective.

In Appendix L the overall summary statistics are given for the output measures and copies of the examination papers are included. The output assessment of the "attitudinal" objectives V(C), VI(A), VI(B) and VI(C) was through a questionnaire to all students. A copy is given in Appendix L.2. The objectives primarily assessed by the supplementary questions in the 73.101 examination paper are also listed in Appendix L.1.

All questions used for the "input" assessment were examined for content validity in terms of the seventh form prescriptions by teachers involved in the pre-testing programme. All questions used for "output" assessment were examined for content validity by the two other lecturers involved in the first year physics courses. Further, careful consideration was given to whether or not the questions had validity with respect to the objectives they were intended to measure.

An external judge to assess the questions used for "output" assessment should have a clear knowledge of the intention of the objectives, should be familiar with the learning activities of the course, and should not have been involved in the development of the objectives, the teaching or the test questions. In a small educational system such a person is not easily found. Fortunately the 1973 Teachers' Fellow in Physics had been closely associated with the first year physics teaching in 1973, had a clear understanding of teaching in schools and was familiar with the first year physics objectives

---

[10] The 1973 final examination was kept secure so that output questions remained virtually the same for both years.

(Nicholas, 1973). At the end of 1973 he attempted on first sight to categorize the "input" and "output" assessment questions in terms of the objective each question was attempting to assess. Of the 112 multiple-choice questions, he correctly assigned 88 (79%) to the intended objective, and a further 17 (14%) were assigned to the correct complex of objectives although not to the specific objective (e.g., II(A) instead of II(B)). Of the Likert items, only one was assigned to an incorrect objective. All free format questions were assigned to the correct objectives. After discussion, and consideration of the pre-test statistics of the questions, the Fellow agreed with the categorization of all but one multiple-choice questions.

This one study did establish that questions could be relatively unambiguously classified in terms of the largely content-independent objectives of the course. In general, the level of specificity of objectives was justified in that objectives were distinguishable in terms of assessment. However, of the 17 questions correctly classified within the right complex of objectives (e.g., III(A), III(B) and III(C) form a complex) but not to the right objective within the complex, 10 of these occurred in distinguishing between II(A), II(B) and II(C) objectives. This suggests that the distinction between these levels is not particularly clear in practice and, in fact, little distinction is made between these levels in teaching. Further validation of the question objections by independent judges would have been desirable but it requires a clear appreciation of the objectives and, as Kropp and Stoker (1966) point out, it is essential that judges have a clear knowledge of the learning experience of the students. Such people were not available.

8.3 The development of attitudinal measures

As outlined in Section 5.3, the scientific "attitudes" of open-mindedness, self-criticism, willingness to suspend judgement, and commitment to accuracy are considered to be a reflection on the nature of scientific inquiry and to be inherent in the higher cognitive objectives (e.g., Objective III(A) and III(B)). Hence the term attitudinal objectives has been restricted to objectives which are characterized by an emotional content and for which there are no right or wrong answers, as occurs in the cognitive area. This then restricts the attitudinal objectives of Table 5.1 to Objectives VI(A), VI(B), VI(C), and to some extent Objective V(C).

Although the semantic differential technique was also investigated, it was decided that Likert scales (Likert, 1932) would be more appropriate for the measurement of attitudinal objectives. In an extensive study of attitudes to physics in Australia, Gardner (1972), after considerable review, used Likert scales exclusively (Gardner, 1972, p.29).

Initially 20 Likert items were written for each of the Objectives V(C), VI(A), VI(B) and VI(C). There was an equal number of positive and negative items. The items were checked for face validity by obtaining responses on the items from senior students and staff committed to physics. The items were pre-tested in two groups, each with 100 students at the end of their seventh form year (November, 1971). Each item was correlated with the total student score for the 20 items on that particular objective.[11] From these results a 40 item test was developed with 10 items on each objective. The items were chosen on face validity, discrimination index[12] and response level. Any item which was too similar to one already chosen was excluded and an equal number of positive and negative items was chosen.

---

[11] Responses to positive items were ranked 5-1 from strongly agree to strongly disagree. Conversely negative items 1-5.

[12] Calculated by using the total score on all items intended to measure the objective under consideration.

This 40 item test was pre-tested with first year physics students in 1972 at the beginning and at the end of the courses. A principal component analysis of the items to check on construct validity led to a reduction to 9 items on each objective.

This finalized 36 item test (see Appendix K, or L) was used with first year students in physics in March and October 1973, and in March and October 1974. In addition the test was used with a random sample of 193 seventh form students in secondary schools at the end of the 1974 school year.

All these data were analyzed set-by-set using a principal components analysis and a varimax rotation. The 5 sets of data yielded similar factor loadings which were, in general, as expected. Objective VI(B) items and Objective VI(A) items loaded heavily on Factor I and Factor II respectively. Objectives V(C) and VI(C) were more complex, loading on Factor III or IV for Objective V(C), and loading on Factor I or III for Objective VI(C). These loading patterns seemed to be compatible with the nature of the objectives. Table 8.1 gives a single analysis of all the data in the 5 sets. Table 8.2 gives the typical factor loadings for the school analysis (October 1974), for an input university analysis (March 1973) and for an output university analysis (October 1974).

In Table 8.1, the two initial items have low communality while the penultimate item has an unexpected factor loading. These loadings were unanticipated from the 1972 analysis of the 40 item test. It would appear that if certain items are placed at the beginning or the end of the test they are likely to operate inconsistently compared with pre-test predictions. This suggests that it is desirable to include dummy items at either end of a test to buffer the test items. Consequently, in determining a student's total score on each objective, items 1, 2, 35 and 36 were not included, and a score on the eight remaining items for each objective was obtained.

In addition to the above analyses, Tennent (1972) investigated the use of the semantic differential as an alternative method of attitude measurement for the first year physics objectives. This work is detailed in Appendix M. It supports the validity of the Likert scales to some extent, although it also indicates some of the complexities involved in measuring attitudes.

8.4 A statistical investigation of student attainment

As stated previously, both input and output tests were designed to measure each of the stated objectives of Table 5.1 although the input tests did not measure Objectives I(A), IV(D) and VII(A) as these did not correspond to any seventh form objective. Wherever possible multiple-choice questions were used so that a variety of content areas could be used to measure any one objective. This was essential for the content-based objectives (I - III). It was considered more appropriate to measure most experimental type objectives by means of short questions with free format answers because questions testing these objectives required students to "invent" ideas, design plans of action, and draw conclusions. This could be done only very poorly by having students select from given alternatives. Attitudinal objectives were assessed by Likert scales as these were considered most appropriate.

Because of the constraints on testing time it was possible to have only a small number of items covering one objective (e.g., 5 multiple-choice questions or 2 short questions or 8 Likert items). It was appreciated that in any model of the system (see PART C) any "input" or "output" measure of any reasonable reliability would need to be a combination of a number of objectives. Objectives could be either combined on the basis of the established philosophical groups (complexes of objectives) of Table 5.1 or combined on the basis of a statistical analysis of the correlations of scores on each objective.[13] A component analysis of the correlations of output objective scores was expected to be more interpretable than an analysis of the input scores. This was because the students' learning experiences were better understood in terms of the output measures, and for this reason questions were confidently designated to be measuring a particular objective (i.e., II(B) or II(C)). Also component analysis of the input scores was likely to be confounded by different teaching styles.

The principal components analyses of the output scores[14] in 1973 and 1974, yielded four significant and interpretable factors[15] with eigen values $\geqslant 1.3$ for each year. These four factors accounted for approximately 54% of the variance in both years. This low percentage was expected, because of the unreliability of individual scores. A Varimax rotation of the four factors yielded similar factor loadings in 1973 and 1974 (see Table 8.3). A consideration of the four factors showed that there was;

I a comprehension and analysis factor (primarily Objectives I and III),

II an application factor (primarily algebraic/numeric problems Objectives II(A), II(B), IV(E) and VII(A)),

III an experimental factor (primarily Objective IV),

IV an attitudinal factor (primarily Objective VI)

It may be justifiably argued that these factors can be viewed as grouping objectives on question type rather than on the fundamental

---

[13] A score on an objective is the total mark obtained by a student on any one objective, e.g., IV(A). It is not to be confused with objective type (e.g., multiple-choice) questions.

[14] In these analyses Objectives V(A) and V(B) were summed before analysis. Some questions measuring these objectives were placed at the end of a section of the examination paper resulting in scores in each objective being too unreliable for individual consideration.

[15] Other rotations were made with three and five factors, but these were not so easily interpretable.

similarities of the objectives. This problem is unavoidable once the decision is made that certain objectives have to be assessed with different types of questions. However, the factor loadings on Objective VI(A) and the two factors for multiple-choice questions must be interpreted in terms other than question type. Moreover, the four factors isolated here are not inconsistent with other component analyses made of tests consisting of only short questions and free format answers during 1973 and 1974.[16]

The above does not contradict the fact, however, that the use of different types of questions has enhanced similarities and differences among objectives. Nevertheless, the four main groupings of objectives identified by a purely statistical analysis are interpretable even if they do not clearly support the purely philosophical groupings of Table 5.1. The above analyses suggest that in terms of the measurement of course objectives in this system, with these types of questions and this type of open book examination, there is no clear statistical distinction among any of the I(B), I(C), III(A), III(B) and III(C) objectives. These objectives relate to a clear understanding of theoretical physics concepts. On the other hand, those objectives emphasizing the intellectual skills of application (solving applied-type problems not necessarily requiring a deep understanding of physics) tend to group together statistically, Objectives I(A), II(A), II(B), IV(E) and VII(A). The abilities closely associated with planning, recording and drawing experimental conclusions undoubtedly involve a certain knowledge of

---

[16] In the analyses of experimental tests it was apparent that Objective IV(E) did not have similar factor loadings to other Objective IV questions, even when all objectives were measured by the same type of question. In the analyses of theoretical test results it was common to find inconsistent factor loadings from one test to another, particularly for Objectives I and III, suggesting that these are complex objectives dependent on subject content and that they are difficult to assess without a number of questions covering the content area. Principal components analyses and other factor analyses tended to group objectives in similar ways and, for simplicity, principal components analyses have been used in this thesis.

physics. However, the scores on questions covering these objectives are statistically clearly distinct from the students' scores on the other objectives. Finally, the scores on the "attitudinal" objectives (except VI(A)) are relatively independent of the scores on the other objectives.

The results of the components analyses were not anticipated from the Table 5.1 grouping of objectives developed for use in the teaching and writing of test questions. Rather the results suggest a different grouping of cognitive objectives which, in retrospect, emphasize three distinguishable aspects of physics;

(i) theoretical understanding,
(ii) application ability, and
(iii) experimental ability.

Again, in retrospect, these three categories are characteristic of the different interests and abilities of three hypothetical "types" of students. The "<u>theoretical</u>" student is interested in fully understanding the concepts and implications of theory, the "<u>applied</u>" student is more concerned with being able to find the right equations and manipulate them to solve the specific problem at hand, and the "<u>experimental</u>" student is often of a more "practical" outlook and is more at home in the laboratory setting. While individual results, or student profiles on the seven "educational" objectives of Table 5.1 would be unlikely to be of interest to another educational course, institution or prospective employer, a reliable profile of a student on such "objectives" as theoretical physics, applied physics and experimental physics (in addition to a comment on a student's industry and attitude to the subject) might well be of greater interest than a single physics mark.

Further analyses are required to establish the generality of the similarities and distinctions between attainments on questions measuring different objectives. The input tests provided an opportunity to see how stable the factor loadings were with different questions and when measuring the same students at the beginning rather than at the end of the academic year. Again accepting eigen values $\geqslant 1.3$, four factors were obtained in a component analysis of the input tests in 1973 and 1974 and these factors accounted for approximately the same percentage variance (52%) as the output tests. The factor loadings on these tests (see Table 8.4) again established;

I a comprehension and analysis factor,
II a problem-solving factor,
III an experimental factor,
IV an attitudinal factor.

A consideration of Table 8.5 and Table 8.6 shows many similarities although "objectives" II(C), IV(E), V and VI(A) are loading on different factors in the input tests than in the output tests.

As a final analysis, the input tests used at the beginning of 1974 were used as seventh form tests at the end of the 1974 academic year in 17 schools (N = 213). The tests were administered by teachers in their classrooms in types of schools (single sex/co-educational) and types of settings (rural/urban). Using the eigen value $\geqslant 1.3$ criterion only three "significant" factors were obtained, and when the analysis (see Table 8.5) is compared with Table 8.3 and 8.4 it is clear that no statistical distinction exists in this case between "comprehension and understanding" and "problem-solving" objectives. The three factors which account for only 44% of the variance are interpretable as;

I a theory-application factor (comprehension, problem-solving, analysis),
II an experimental factor,

III an attitudinal factor.[17]

In the input analyses (Table 8.4) the statistical distinction between objectives loading on factors I and II was not particularly clear. In the school situation the students experienced a variety of learning situations immediately prior to the test (depending on the learning programme in each school) and the test was administered in different ways (some used parts of the tests for formal assessment, others did not). Thus it is not surprising that some of the distinctions clear in the analyses of Tables 8.3 and 8.4 were submerged. Nevertheless, there are similarities of factor loadings in Tables 8.3 and 8.5 despite the fact that they apply to different tests, different students and a different academic year.

Although the components analyses of the input and output tests did indicate four reasonably stable and interpretable factors, other components analyses of first year cognitive tests between 1972 and 1975 confirmed that correlations between questions, and hence the factors and factor loadings obtained, are influenced by question type, content area of question, position in the test of the question, the number of different questions in the test, and the objectives being assessed. In the analyses of Tables 8.3 and 8.4 (and 8.5) every objective was measured by more than one question to decrease the effect of question position and content area, and no unnecessary variation of question type was introduced. In the output tests all students had experienced basically the same learning system and teachers. Even at the input test stage some of the individual learning differences due to different schooling may have been lessened by the common October examination and

---

[17] Removal of the $\geqslant 1.3$ eigen value criterion and the inclusion of the fourth eigen value ($\lambda_4 = 1.2$) failed to produce loadings on the four factors which could easily be related to those of Table 8.3 and Table 8.4.

the 3 month break from schooling which all students had experienced prior to the test. In terms of assessing students from a variety of schools and hence a variety of learning experiences, the output from schools analysis (Table 8.5), when compared with earlier analyses, suggest instability in the factor loadings on Objectives I and II. In this area it is difficult to specify the objectives being assessed by any one question without knowledge of the learning experience and the test situation. Components analysis of national examinations designed to measure aspects of objectives (e.g., see Appendix I) support this point of view.

The importance of the number and the type of the variables in the components analyses cannot be overstressed. If only attitudinal objectives, for example, are used in a components analysis, the differences between these objectives will be emphasized (see Table 8.2). If the attitudinal objectives are analyzed along with cognitive objectives, the similarity of the attitudinal objectives, compared with the other objectives, may well be emphasized (see Table 8.3). Although in an analysis involving all the objectives of first year physics, the experimental objectives tend to load on the same factor, components analyses of only experimental objectives questions tend to show differences among these objectives. In particular, Objective IV(D), Objectives IV(A) and IV(B), and Objectives IV(C) and IV(F) often loaded on different factors when tests were analyzed. One consistent pattern observed in all test analyses was that questions measuring Objective IV(D) always loaded on the same factor as the laboratory book assessment mark. This assessed mark on the laboratory book is allocated on the basis of the completeness of a student's laboratory book—an indication of the student's industry at least in the experimental course.

In summary, it is difficult to measure reliably all the cognitive objectives of the courses not only because of the limited testing time available but also because the valid measure of an objective depends upon the learning experiences of the students being well known. Even in the case of the output assessment, where the learning experiences of the students were well known, the results in Table 8.3 did not support the Bloom-based categorization of the first year physics objectives given in Table 5.1. Rather four main objective areas were identifiable, namely, "theoretical", "applied", "experimental", and "attitudinal". It must be emphasized, however, that the similarities and differences among objectives were undoubtedly enhanced to some extent by the type of questions used to assess the objectives. Nevertheless, the groupings established in Table 8.3 appear to have potential for the preparation of meaningful student profiles interpretable by interested people not closely involved with the courses.

CHAPTER 9

THE MEASUREMENT OF COURSE EFFECTIVENESS (PRODUCT EVALUATION)

OUTLINE: Consideration is given to the standard of attainment of students with respect to the cognitive objectives, and to the change in attitudes of students with respect to attitudinal objectives. Students' perceptions of course effectiveness are briefly outlined. Finally, the problems of adequate product evaluation are discussed.

9.1 Cognitive objectives

As pointed out in Chapter 8.1, the measurement of course effectiveness in absolute terms is not possible where the objectives of the course relate to general learner objectives. However, these general objectives were useful in other ways, and the problems associated with other types of objectives and their measurement resulted in their retention and use for the first year courses. It can be argued that although product evaluation is concerned with the achievement of the course objectives, course effectiveness is a much broader issue and involves the appropriateness of the objectives (context evaluation) and long-term consideration of the effect of the courses on students. This is a valid view. However, these wider issues will not be considered here. Course "effectiveness" is here considered primarily in terms of product evaluation, and hence limited to the attainment of the stated objectives.

With regard to cognitive measures and product evaluation, test and examination questions tended to be both criteria and normative based in that questions were developed by academic staff on subjective criteria moderated by knowledge of the previous performance of students.

Initially, in 1970 for example, without previous knowledge of student performance and experience of open book examining, there was a tendency for examiners to expect too much of students. Also students were given insufficient guidance on how to cope with an open textbook examination. Subsequently, staff expectations, particularly with respect to time allocation, were revised and the limited advantages of the open textbook examination for many objectives was pointed out to students. Although rather informal, feedback obtained from marking tests and examinations and from the associated statistical analyses of results is an important aspect of product evaluation, and this feedback significantly affected the teaching, testing and the objectives themselves, particularly in the early stages of course development.

In the development of the output tests for 1973 (and 1974) from pre-tested items, the difficulty of obtaining "suitable" questions (in terms of facility in particular) for some objectives suggested that either the objectives were unrealistic or the teaching inadequate. For example, it was particularly difficult to find questions on Objective II(C) which genuinely involved a new application situation and yet for which the facility level was sufficiently high to discriminate among students ($F > 0.3$). This suggests that students have difficulty applying the knowledge to new situations. In terms of context evaluation the appropriateness of Objective II(C) needs consideration. In terms of input evaluation consideration needs to be given as to how best to teach the objective. Pre-tested Objective III(A) questions were also generally of a very low facility level. In this case a low facility was expected but again findings need to be related to context and input evaluation. On the other hand, Objectives I(A) and VII(A) questions were difficult to find with a sufficiently low facility level although to some extent high facility levels were appropriate for these objectives. For some objectives it was particularly easy to find questions which were considered to discriminate on that objective and yet which were of a reasonable facility level, for example, Objectives I(B), II(A), II(B) and IV(E). This suggests that for these objectives there was a good match of objectives, teaching, and assessment.

For many objectives, the questions designed to measure that objective tended to be of similar facility. However, for some objectives the facility level was difficult to predict for untested items and was prone to considerable variation with slight question modification. These objectives were I(C), III(B) and III(C) and the variation suggests that these objectives are particularly responsive to learning experiences and little influenced by the normal distribution of aptitudes (mathematical aptitudes, for example). Objective I(C) directly relates to teaching emphasis in lectures and laboratory and, in retrospect, Objectives III(B) and III(C) were given specific emphasis in the learning programme.

The reconsideration, after the course has been developed to a reasonably stable state, of the facility levels of the pre-tested items used for product evaluation can, on its own, be expected to add little new knowledge about course effectiveness. In 1973 and 1974 the difficulty level of questions was considered appropriate by all academic staff involved in the courses and the overall examination results required little adjustment to provide "comparable and reasonable" pass rates compared with other subjects within the university.

9.2 Attitudinal objectives

In the first year courses, student assessment for the purpose of a final grade was not the only assessment undertaken. Primarily because of the interest in an attempt to develop statistical models of the system (Part C) attitudinal objectives were also assessed even though these did not form part of the official student assessment. As security was not a problem here the opportunity was taken to use the same test prior to, and after instruction.

Attitudes were not expected to change so significantly that the input test would be inappropriate as an output test. In terms of product evaluation, some of the serious problems of a pre-test and post-test method of evaluation have been detailed (Section 8.1). Using the same test prior to, and after instruction provided an opportunity to reconsider these problems.

Using the finalized attitude test, described in Section 8.3, the results shown in Table 9.1 were obtained. Both the input and output means exclude the 9 "dropouts" who did not complete the course in the years 1973 and 1974.[1] The overall scale means were high.[2] However, the means were generally lower by the end of the year albeit not significantly in most cases. The significant drop in the mean for Objective VI(B) needs further consideration. Despite these trends, student replies to the explicit question, "how have your attitudes changed from this time last year?" (see Table 9.2) showed that only 15-17% of students considered their attitude was more negative. On the contrary, approximately 50% of students considered their attitude to be more positive.[3]

"Input attitudes" were a retrospective measure in March of student attitude toward their seventh form course at the end of the preceding year. Unfortunately, it was not possible to measure their attitudes directly in the preceding October as the students were unknown at that time. If this had been possible it may have removed a likely systematic

---

[1] In 1974 four students were included who did not sit the final examination.

[2] Scale category weightings were: Strongly positive ~ 5, positive ~ 4, neutral ~ 3, negative ~ 2, strongly negative ~ 1.

[3] Students could remain anonymous if they so wished so that replies should not have been unduly influenced by any desire to please.

error introduced by not measuring their attitudes at the same time in both years. At the beginning of the year students have an enthusiasm which is singularly lacking in educational institutions just prior to final examinations, when the output score was obtained. As a check on this possibility, the mean score of a random set of 214 seventh formers on the attitudinal test was obtained in October 1974 to coincide with the time of the year at which the output measure was made in the university. While some of the students with poorer attitudes to physics may not proceed to university physics, and this will deflate the school scores somewhat, the results (see Table 9.3) do support the possibility of a systematic error affecting the absolute change in attitude as measured by the input and output tests. A decrease in measured attitudinal scores over a course of instruction is not uncommon (for example, see Gardner (1973), Alexander (1974) and Harlen (1975)).

The presence of a systematic error does not necessarily affect the comparison of individual student scores. The correlation of stated attitude change with measured attitude change is given in Table 9.4.[4] These correlations suggest that students view "attitude" primarily in terms of Objective VI(B). The correlation is significant ($p < 0.01$) but not large.

Further, the mean output score for each attitudinal objective was obtained for three different groups of students, namely, those students who in the subsequent year intended;

(i) to study at least some second year physics,

(ii) to specialize, i.e., in engineering, medicine, and so on,

(iii) to discontinue a study of physics beyond first year.

---

[4] With respect to the questionnaire in Appendix L by which students stated attitude change was obtained, a linear marking of 5 to 1 was placed on options A to E of the attitude change question (PART C, Q.1).

Table 9.5 suggests a link between attitude and the intention to do more physics, particularly with respect to Objective VI(C). This supports the validity of the VI(C) score.

Unfortunately, after all these considerations are taken into account, one is still left with little clear indication of how successful the courses have been with regard to the attitudinal objectives. As an alternative evaluation procedure, a comparison of the product of one system with a similar system may have been useful, but this requires a system not only with similar students but also with similar objectives. Gardner (1973) investigated attitudes of sixth formers to physics in Australia. While there are considerable differences between the system investigated by Gardner and the first year physics courses at Waikato, the nature of some of the objectives and the method by which they were measured were similar. Gardner also found a decrease in attitude means similar to that in this study.

Yet another alternative is to consider the somewhat hierarchical nature of the objectives (see Krathwohl, 1964) and consider alternative ways of measuring the "highest" attitudinal objective, Objective VI(C), "developing a commitment to physics". Objective VI(C) is related to the retention rate of students from first year in more advanced physics courses. Compared with the national average for the retention from first year physics students in physics majors the Waikato figures look healthy (Table 9.6) but again the intentions of entrant physics students from one university to another differ. Of course, students are also influenced in their decision to advance in physics by factors other than their "commitment to physics" attitude.

In summary, while explicit statements of attitude changes by students, and satisfactory retention rates in higher level courses in physics, are indicative of positive student attitudes, the more direct attempt to measure student attitude to physics using Likert questions for a pre-test and post-test, produced disappointing results. These may have been due at least in part, to a ceiling effect in that overall attitudes were positive, and also due to a systematic error introduced by measuring input attitudes in retrospect. However, it would appear pertinent to reconsider both the assessment procedures and the emphasis in teaching with respect to these attitudinal objectives.

9.3 Student view of course effectiveness

Student questionnaires emphasized and yielded student responses related to process evaluation. In addition some free responses by students were concerned with more general issues and hence more related to product evaluation and overall course effectiveness. However, rarely did a student make a comment on the aims of the course, and whether or not these aims were being achieved. In retrospect questions could well have been designed to elicit responses from students about;

(i) the relative importance of the explicit objectives, and also

(ii) whether or not the course effectively developed each objective.

Many students commented that the courses were informative. No student stated the converse! Some felt that the physics concepts, although not difficult in themselves tended to be obscured by the mathematical language used. With regard to course content, it was frequently stated that there was too much material in the courses (particularly the theoretical course) although there were some more positive comments. For example, "I think I have learnt more physics this year, than in three years of school physics put together. I have appreciated the laws better and come to a better understanding of them".

Very few students made comments which were pertinent to the explicitly stated course objectives. A few students considered the courses could have provided more technological applications and some felt insufficient attention was given to "where physics is heading". This relates to Objective V(C).

A common view of students not continuing with physics was that they enjoyed physics or found it interesting but they had insufficient confidence in their mathematical or physics ability to feel that they could be certain of success in advanced physics courses. This lack of confidence was in most cases not misplaced, and indeed, a function of student assessment is to provide them with data on which they can base their decisions regarding their future education. As expressed flamboyantly by one student (October 1972), "I am irrevocably persuaded that physics without a mathematical brain is an incongruity."

Comments on the techniques used for student assessment showed that most students prefer the open book examination system. The completely open book tests and examinations in the experimental course appear to have been particularly successful.[5] A rushed and rather colloquial free response to the October 1974 survey was "The tests are good 'cos [sic] you just work your way through them without getting in a flap about things you cannot understand properly. In fact I quite enjoy 73.102 tests although in all other courses tests tend to be a hassle. I learnt a lot from 73.102 tests even though I have not managed to pass one yet." Although many students would endorse this positive approach, early tests were often considered by students somewhat long and requiring fast working techniques. Not one free format comment was made in five years on the objectives being assessed by the tests. As to student confidence in the marking of tests there were only two comments on this topic. These can be lightheartedly placed in juxtaposition. Both replies were

---

[5] From a teaching point of view the tests and examination in the experimental course appeared to give the course objectives the correct emphasis. The emphasis that students placed on different aspects of the course appeared to be influenced by the proportion of examination questions on particular objectives. By changing the proportion, the aspects of the course emphasized by students could be altered.

answers to the question about whether they had learnt more from the experimental or from the theoretical course. The comments were:

(i) "My mark gained in 73.102 tests is higher than the average gained in 73.101. Therefore I have learnt more in 73.102."

(ii) "I have learnt more in 73.102 but get lousier marks in this."

While there has not been an increase in negative comments on the courses in recent years, there have been fewer free responses of a very positive nature about the overall informational and interest value of the courses by students. Possible reasons for this are;

(i) that the lack of negative critical comment has led to a greater emphasis in recent years on eliciting criticism from students which point out weaknesses rather than give general praise,

(ii) that in the first few years the academic staff, somewhat unsure of the new courses, unconsciously felt a need for ego-bolstering comment and sensitive students felt they should provide it(!),

(iii) that some of the new ideas initially used in the course have now become more widely used, and the equipment and textbook are obviously no longer new,

(iv) that students' views towards physics are changing.

9.4 Summary of product evaluation

In the first year courses, the course objectives were specified in terms of student behaviour. Hence product evaluation was associated with student assessment. The level of specificity of the objectives considered appropriate for the courses meant that product evaluation was necessarily subjective. However, feedback from test and examination answers to questions specifically orientated to particular objectives was useful for understanding what was being achieved, what one could reasonably expect from students, whether or not objectives (as elucidated through questions) were reasonable goals to expect students to achieve, and whether or not teaching methods with respect to particular objectives needed further consideration.

In addition, the more careful development of assessment measures in order to attempt to assess all objectives adequately in 1973 and 1974, appeared to have had a beneficial effect on teaching and on students' efforts to achieve at least the cognitive objectives.[6] A statistical analyses of the relationships between student performance on all objectives yielded a different categorization, or grouping, of objectives compared with the original groupings (Table 5.1) which were based on Bloom's educational classification (Klopfer, 1971). While it is clear that groupings are somewhat dependent on the decisions made about how best to assess particular objectives, statistical analyses may well lead to more relevant and valid groupings of objectives for particular learning systems and particular subjects. For example, the objectives of Table 5.1 can be regrouped as shown in Table 9.7 (on the basis of Table 8.3 and the supporting evidence of Section 8.4). These groups can be interpreted as representing three areas of physics activity—theoretical, applied and experimental physics—while a fourth area is the appreciation of physics. The statistical evidence given in Section 8.4 for these groupings is not particularly strong and only indicates that within these courses, measured by these assessment procedures, this new grouping of objectives may be more appropriate.

---

[6] One might conjecture that if the objectives and the assessment procedures of a university course are appropriate and satisfactory respectively, then it is likely that adequate learning will follow. Conversely, if objectives are stated but no attempt is made to assess them, little emphasis will be given to them by students. Considerable student frustration occurs when teaching emphasis is towards unassessed objectives.

However, from a philosophical point of view the new groupings are not unreasonable and, in fact, may well be more acceptable, meaningful and useful for physics teachers and students than the "educationally based" groupings of Table 5.1.[7]

In the long-term more relevant and valid groupings of objectives in particular systems could lead to more appropriate methods of discrimination analysis, and to a better understanding of product evaluation. However, there are difficulties in obtaining reliable measures of achievement for particular objectives. As suggested in Section 8.4 it is increasingly difficult, as the learning experiences of the students become less well known, to classify questions in terms of particular objectives.

In conclusion, product evaluation is seen as an important aspect of evaluation which needs further development. In the first year physics courses objectives, teaching, and assessment were developed to the point where questions could be devised for all objectives at a level of difficulty which those teaching the courses considered appropriate and which, at the same time, provided pass rates considered appropriate by the university without significant scaling. It is difficult to quantify adequately this attainment level, and even more difficult to quantify the extent to which students' attainment level was changed by the instructional programme. In terms of the total educational

---

[7] Table 8.3 not only suggests possible changes in the grouping of objectives but it also provides evidence which supports other changes which may appear desirable for a variety of reasons;

(a) in terms of testing: Objectives I(A), V(B), VI(A) to VI(C) could well be assessed at a more sophisticated level,

(b) in terms of teaching: Objectives V(A) to V(C) deserve greater emphasis,

(c) in terms of objectives: The distinction between Objectives II(A) and II(B), between Objectives II(C) and III(A), and between Objectives I(B), III(A) and III(B) may not be useful and alternative distinctions among objectives in these areas may be more appropriate.

system (see Figure 4.2) it may well be fruitful from a product evaluation point of view to obtain the views of students about the courses some years after instruction. In this way the long-term effects of a course could be assessed albeit again subjectively. In the short-term, the development of techniques to assess student achievement of stated objectives, and the feedback from this for the purpose of further refining and understanding assessment procedures, course objectives and teaching methods appear to be the best way of improving product evaluation and hence learning.

PART C

THE DERIVATION OF SPECIFIC SYSTEMS MODELS

CHAPTER 10

THE DEVELOPMENT OF STATISTICAL MODELS

OUTLINE: In Part C consideration is given to the "basic" aspects of the systems approach and specifically to the development of models of the first year physics courses. This chapter provides an introduction to this work. Consideration is given to the variables pertinent to the system and the statistical techniques to be used in the models.

10.1 The variables pertinent to the system

The majority of students taking the first year physics courses had previously undertaken a seventh form year of physics study. The learning experiences of this group were relatively uniform and the first year physics courses were designed basically for students with this background. Minority groups of students had different backgrounds; students with four years secondary schooling, students with five years schooling without seventh form physics, overseas students, older students, repeat students and so on. However, none of these groups was sufficiently large to be treated as a separate statistical model. For these reasons, as well as a number of minor administrative reasons,[1] a decision was made to limit the model-building study to students who had experienced a seventh form year of physics. This resulted in the inclusion of the majority of students, and in a homogenous group with regard to academic background.

The development of any model of an educational, or even a physical system, is dependent on the variables considered to be of importance and

---

[1] For example, many overseas students enrolled late making input testing difficult.

to be investigated. This, in turn, depends on the purpose of the model. In this study the purpose of the model was to elucidate the teaching system, and hence, in the long-term, to provide knowledge and feedback useful for decision-making with respect to objectives, teaching methods and assessment within the first year physics courses.

In practice there is a limit to the number of variables that can be investigated. Crosslund and Moore (1974) give the following reasons for excluding variables:

(i) The time and effort involved in collecting data on a particular variable may not be warranted in view of its marginal effect.

(ii) The time and effort involved by others in providing the data on a particular variable (e.g., student or former teacher) may be impractical or unreasonable.

(iii) The collection of data on a particular variable may cause excessive embarrassment or arouse unwelcome hostility and perturb the system.

(iv) The data collected may be inaccurate or unreliable.

Two major factors influenced the choice of variables considered in this research. First, Miller (1970), in a survey of studies investigating academic success, pointed out that the findings about what background and environmental factors were important in academic performance were varied and somewhat inconsistent and that the important factors were dependent on the specific system under study and the particular variables chosen for inclusion. Secondly, studies of the developing first year physics courses in 1970-1972 showed that there was a high correlation between seventh form physics achievement and final achievement in the first year physics courses ($r > 0.7$, see also Table 10.1 for the 1973 and 1974 correlations). To provide useful information for elucidation of the objectives-learning-assessment system, emphasis on the components of earlier academic achievement rather than on personality or psychological background factors seemed desirable.[2]

Other evidence seemed to support this view. Statistical investigations of an exploratory nature in 1970-1972 with the first year physics students suggested that variables such as age, birth rank number in family, were not significantly correlated with physics achievement, while a student's aspirations with respect to the course and the parents' level of education were only of possible significance. Further, at the end of the first year courses in October 1971 a small group of students (N = 23) were asked to consider a limited list of environmental and background factors which may have influenced their achievement in physics either advantageously or otherwise during the year. These students were asked to list the five most important factors. In terms of frequency of inclusion the most important factor was "background in physics". Of secondary importance were "background in mathematics", "interest in physics", "motivation to do well" and "academic demands of other subjects". Of little importance were "intellectual ability", "friends' interests", "enthusiasm of laboratory partner", and academic or emotional disruptions. Students did not consider that any factors of importance had been omitted from the list.

In view of the above, prime importance was given to obtaining a detailed knowledge of students' background achievement in physics not only by developing a set of input tests related specifically to the course objectives but also by obtaining from the Universities Grants Committee full details of student performance in the seventh form

---

[2] Conversely, if there had been a low correlation between input and output achievement it would have been of greater interest and likely profit to place the emphasis on isolating and measuring non-academic background factors.

Bursary physics examinations.[3] In addition, students' mathematics achievement in the Bursary mathematics examination was obtained along with academic performance in other school subjects.

To give some indication of students' basic verbal and mathematical ability the students' School Certificate results in English and Mathematics respectively were also obtained. These results related to their school achievement 2½ years prior to entering university. To obtain a measure of a student's aptitude which may affect achievement in physics, but which is neither dependent on earlier schooling nor related closely to mathematical or verbal ability, modified versions of the abstract reasoning and spatial relations aptitude tests (DAT, 1963) were developed. These shorter tests, 17 min and 21 min respectively, were administered with the "input tests" (see Appendix K.3) at the commencement of the first year courses.

Each student's former seventh form physics teacher was asked to assess the student's academic industry in the seventh form on a seven-point scale along with a similar assessment of the student's independence and stability of purpose. It was considered that these two assessments together would give an indication of a student's motivation for academic work.[4]

In addition to the above information, data was also obtained about each student's stated aim in each course in terms of his anticipated grade and the highest educational attainment of either parent (see Appendix K.5). A variety of input physics achievement

---

[3] This was obtained with special permission of the Secretary of the Universities Entrance Board and the Chief Examiner of the Bursary physics papers (1972 and 1973). Their co-operation was most appreciated.

[4] Details of the assessment sheet are provided in Appendix N. A number of other scales was also included and these scales will be discussed in detail in Chapter 13.

subtotals were obtained and will be detailed in the appropriate chapter. Table 10.2 summarizes all the "input" variables. The output variables will be detailed in the appropriate chapter.

10.2 Path analysis and the first year physics courses

In terms of the generalized curriculum models proposed in Chapter 4, path analysis provides a technique for developing a statistical model which can be easily related to these more general diagrammatic models. Further, it has the advantage of providing a meaningful model to summarize the intercorrelations among variables in that all the information of the correlation matrix can be retained in the path model. Yet, at the same time, certain simplifying assumptions or postulated relationships can be made and these are testable in terms of the observed data.

As mentioned in Section 4.3, the numbers of students anticipated when this research was planned in 1970-1971 did not materialize in 1973 or 1974. This introduced a problem into the use of path analysis in this particular study. As Cooley and Lohnes (1971) have stated, normalized regression coefficients (these are also the normalized beta weights or path coefficients) tend to be particularly unstable when small samples are used and independent variables are correlated.[5] They suggest that regression factor structure coefficients are more stable and provide a more useful picture of the relationships between

---

[5] As an example, consider that Z_1 and Z_2 in Figure 10.1 are highly correlated and approximately equally correlated with Z_6. In factor analysis of Z_1 to Z_5, Z_1 and Z_2 would tend to load on the same factor irrespective of minor variation in relative size of the other correlations. In a multiple regression analysis of Z_6 with Z_1 to Z_5, however, the relative sizes of the path coefficients p_{61} and p_{62} would be critically dependent on the relative sizes of the Z_{61} and Z_{62} correlations. If $r_{61} > r_{62}$ then $p_{61} >> p_{62}$ and in fact p_{62} may be insignificant. If $r_{62} > r_{61}$ then the converse would apply.

variables than the normalized beta weights.[6] In table 10.3 the normalized beta weights and the factor structure coefficients are provided from typical data obtained in 1973 and 1974 for the first year physics courses. Admittedly the beta coefficients do also depend on correlations not shown. However, there is a much greater similarity in the factor structure coefficients for the two years than for the normalized beta weights. This typifies the problem.

To investigate this further an exploratory multi-stage path analysis model (Type II model—Section 4.2) was developed which included 11 variables. All 41 possible "causal" paths were initially postulated in the model. Of these paths 21 were not statistically significant (5% level) in both 1973 (N = 92) and 1974 (N = 73). Twelve paths were statistically significant in both years. However, there were 8 paths which were statistically significant in 1973 or 1974 but not in both years. The problem at this point was to determine whether the inconsistencies were due to significant changes in the system, in the students, in the measurement of variables between 1973 and 1974, or in the random statistical fluctuations in the basic correlation matrices. While all these were likely to have had some effect, there were no really significant changes in the students or the measurement of variables from 1973 to 1974 (e.g., see Table 8.3). In addition, interest in this research was on the instructional system and this was considered not to have changed significantly from 1973 to 1974.

In view of the above, it was decided to make the simplifying assumption that the educational system was basically unchanged from

---

[6] Regression factor structure coefficients give the correlation of an independent variable to the regression function. It is obtained by dividing the predictor-criterion correlation by the multiple correlation coefficient.

1973 to 1974 and that the most useful model, suggested by the data, would be a path analysis "mean" model for the two years. Correlations for all variables, used in this "mean" model, would be first obtained for each year separately and then transformed to Fisher's Z coefficients (Guilford, 1965, p.589). Weighted Z coefficient averages would be obtained by taking the number of students in each year into account and hence the average correlation coefficients for the two years obtained.

While it may have been preferable to analyze each year separately in that the second year could provide cross-validation, unreliable path coefficients simply generate spurious models. Alternatively it could be argued that path analysis was not appropriate for this study. However, the problem with factor structure coefficients for example, is that much of the original correlation information tends to be lost and it is difficult to gain a clear understanding of the total interrelationships between the variables from these coefficients alone. Similarly other statistical procedures such as discriminant analysis are not appropriate for multistage models. In addition there are advantages in using path analysis in terms of the underlying generalized curriculum model, the potential of the method for the study of educational systems, and the retention of all information about the system by the model. However, in addition to the path analysis models developed in Chapters 11-13, the factor structure coefficients have normally also been recorded.

10.3 Input-output and multistage models

The simplest input-output model is the multiple input-multiple output model of Figure 10.2(a). Consider a particular example of a three input-three output model. In terms of a path analysis diagram, Figure 10.2(a) can be respecified as shown in Figure 10.2(b). In any educational system residuals will be significant. The components of the input and output could be any student attribute. However, for the reasons outlined in Section 10.1, emphasis was given in this research to components of academic achievement.

Three different viewpoints were investigated using this basic model. First, an objectives-model was investigated (Section 11.1). The components of y were student scores on the four groups of objectives identified in Table 8.3, and the components of x were in one model student scores on the four groups of objectives identified in Table 8.4, and in a second model were student total scores in the Bursary physics and Bursary mathematics examinations. Secondly, a content-model was investigated (Section 11.2) where the components of y and x were the student achievements in different content areas of the learning programme (mechanics, fields and waves). Thirdly, a question-type model was investigated (Section 11.3) where the student achievements assessed by different types of examination questions provided the components of x and y.

In the three models outlined above all the input variables for each model were measured at one point in time. This also applied in the case of the output variables. No "causal" connection was assumed, or implied, between any of the input variables or between any of the output variables. If the general model of Figure 10.2(a) is broadened to include other background variables then a multistage model, for example see Figure 10.3(a), may be more appropriate. In terms of a path analysis diagram Figure 10.3(a) can be respecified as shown in Figure 10.3(b).

Causal ordering between variables may be postulated on the basis of time sequence and theoretical assumptions.[7] It was assumed that "a" could possibly affect "b" if "a" occurred before "b" but "b" could not affect "a". However, a problem occurs when variables are measured in retrospect. It was assumed that the retrospective data is the same as the actual data had it been collected in the correct temporal sequence. For example, data on sex, School Certificate and Bursary marks were assumed to be unaffected by retrospective collection.[8] A larger assumption had to be made in assuming that a student's spatial and abstract reasoning aptitudes had not altered significantly from a few years earlier and that a teacher's subjective assessment of a student's industry, and so on, had not altered significantly since the student left the teacher's classroom four or five months earlier. Various multistage models were investigated and these will be detailed in Chapter 12.

Finally, in Part C, Chapter 13 investigates some subsidiary models which consider the problem of feedback, teachers' subjective assessments, and output variables not considered in earlier models.

---

[7] If the order is not clear then various orders may be postulated and each model tested for compatibility with the observed data as Jonnada and Fegley (1974) have done in applying a systems approach to chemical engineering.

[8] School Certificate marks were obtained from two separate sources and cross-checked (the school and the student). Bursary marks were obtained from official Universities Entrance Board records.

CHAPTER 11

INSTRUCTIONAL INPUT-OUTPUT MODELS

OUTLINE: Three input-output models are investigated using multiple linear regression made explicit by the use of path analysis and regression factor structure coefficients. Three different aspects of physics attainment are investigated using these techniques. These aspects are: student attainment of different objectives, student attainment within different content areas, and the effects of using different types of questions to measure attainment.

11.1 The objectives model

In Section 8.4, four aspects of physics attainment were identified. These aspects were isolated by principal components analyses of both the input and output tests designed to measure the relative attainment of students on the objectives of the first year course. The four aspects of physics attainment were;

I a "THEORETICAL" physics ability; as knowledge, comprehension and analysis objectives loaded heavily on this factor,

II an "APPLIED" physics ability; as objectives measuring mathematical physics skills and simple problem-solving ability loaded heavily on this factor,

III an "EXPERIMENTAL" physics ability; as objectives measuring experimental abilities almost all loaded on this factor,

IV an "ATTITUDINAL" physics response; as three objectives assessing attitudes to physics loaded on this factor.

As the groupings of the scores on each objective into these four categories had some statistical support (see Tables 8.3 and 8.4) it was decided to obtain a total score for each student on each of these "factors". This could be done either by using factor scores (a weighting on each objective as suggested by the principal components analysis) or simply by the addition of standardized scores for the objectives loading heavily on each factor. The first procedure was not adopted;

(a) because of the instability of factor loadings from year-to-year caused by the small numbers involved, and

(b) because of the difficulties of interpretation and of communication to others not familiar with factor analysis.

The second method enabled use of the same combination of objectives for both 1973 and 1974 data. The groupings of the objectives established in Tables 8.3 and 8.4 provided four output totals and four input totals respectively. The score on each objective was standardized before addition so each objective contributed equally to the appropriate "complex of objectives" total.

The way the reliabilities are estimated depends on how the scores are to be interpreted.

(i) If the scores are simply assumed to be the scores resulting from the use of particular questions on these particular students on this particular day, then the only reliability to be considered is marker reliability. All multiple-choice results were card punched and verified and subscores on each objective obtained using an IBM 1130 and software developed for the purpose at the University of Waikato computer centre.[1] Total scores were checked against scores obtained by hand marking. For free format answers a marker/re-marker analysis yielded a coefficient of

---

[1] The programme is known as ROMCP and is available from the computer centre.

reliability of $\geqslant 0.98$ for each objective score.[2]

(ii) If the scores are to be interpreted as an attempt to measure an underlying theoretical complex of objectives indicated by the questions used to assess these complexes, but not necessarily these particular questions, then a low limit estimate of reliability can be determined by a split half reliability estimate using correlations of half totals of standardized scores on each objective. This will be very much a low limit estimate because although scores from each objective within a "complex of objectives" tended to load on the same factor (Table 8.3 and Table 8.4) no unequivocal loadings occurred and the objectives within a complex cannot be assumed to be unidimensional.

The coeffficents of reliability of the input and output scores on each "complex of objectives" was assessed using the Spearman Brown formula (Guilford, 1965, p.457) and mean values for 1973 and 1974 are shown in Table 11.1. As these are simply estimates of reliability, they have not been used to attenuate the coefficients of correlation for the path models. Hence the path diagrams should be interpreted as showing simply the relationships between particular subtotals in the case of a particular set of students. The reliability estimates are useful, however, in that they provide an estimate of the correlation of the "complex of objectives" score with an independent measure of the same complex had it been obtained. This estimate can be compared with the correlations of the "complex of objectives" score and the score on any other "complex of objectives". The lower reliability of the input applied

---

[2] Questions were marked using a mark schedule but no marks were recorded on scripts. Later the questions were re-marked by the same examiner with the same mark schedule. A separate analysis using a different marker with his own mark schedule gave an inter-marker coefficient of reliability of $\geqslant 0.90$ for each objective.

score results from the few objectives loading heavily on factor II in Table 8.4.

The four-input to four-output path diagram was postulated to have all possible paths and from the four multiple regression analyses all paths not significant at the 5% level were eliminated and the multiple regression recomputed.[3] Figure 11.1 shows the final path model, and will henceforth be called Model 1. Of the 16 possible paths, nine paths were found to be not significant at the 5% level of probability. In Table 11.2 the measured (observed) correlations are shown below the diagonal in the correlation matrix. The correlation values associated with Model 1 are given above the diagonal. Values underlined once may possibly differ from the observed values and this will be because paths which are not significant have been deleted in Model 1. Values with a double underline may differ from the observed values for two reasons. First, because paths which are not significant have been dropped from the model and secondly, because the chosen independent variables do not totally account for the correlations between the dependent variables, i.e., the residuals are correlated. The residual correlation coefficients are detailed in Model 1 and account for the discrepancy between the observed correlations and the double underlined correlation values associated with the model.[4] In Table 11.3 the associated regression factor structure coefficients and multiple regression coefficients are detailed.

Although no general conclusions are being drawn from the single path model a number of points are worthy of note. The independent

---

[3] All the Chapters 11 to 13 statistical analyses were carried out by the author at the University of London (June-December 1975) on the ULCC-CDC 6000 and also on the Chelsea College, ICL-4130 using SPSS on both machines.

[4] This single and double underline convention will be used in subsequent correlation matrices associated with path analysis models.

variables accounted for about 50% of the variance in the dependent output variables. INPUT THEORETICAL provided the most significant path coefficient not only for OUTPUT THEORETICAL but also for OUTPUT APPLIED and OUTPUT EXPERIMENTAL. The lower reliability of the INPUT APPLIED score and the questionable validity of assessing experimental abilities with hypothetical situations are two factors which undoubtedly had some bearing on these results. However, INPUT EXPERIMENTAL did have a significant path coefficient to OUTPUT EXPERIMENTAL. Also INPUT ATTITUDINAL provided the only significant path to OUTPUT ATTITUDINAL.

BURSARY PHYSICS and BURSARY MATHEMATICS (total scaled scores) were used as the independent variables for a second model with the same outputs as in Model 1. For BURSARY PHYSICS a coefficient of reliability of 0.92 was calculated from the scores of students in this analysis using the split half method and the Spearman-Brown formula. The coefficient of reliability for BURSARY MATHEMATICS was assumed to be comparable. The two-input to four-output path diagram was postulated to have all possible paths and from the four multiple regression analyses all paths not significant at the 5% level were eliminated and the multiple regressions recomputed. Figure 11.2 shows the final path model, and will be labelled Model 2. Of the eight possibly significant paths, three were found to be not significant. Table 11.4 shows the observed and model-associated correlation coefficients using the convention established for Table 11.2. Similarly Table 11.3 shows the factor structure and multiple correlation coefficients.

When Model 2 is compared with Model 1 it is not surprising that the residual path coefficients for OUTPUT ATTITUDINAL in Model 2 is so high (the independent variables accounting for only 6% of the variance). In Model 1 the cognitive inputs made a negligible contribution to OUTPUT ATTITUDINAL.

A comparison of the path coefficients to each of the three cognitive outputs in Model 1 shows a clear distinction between OUTPUT EXPERIMENTAL on the one hand and the OUTPUT THEORETICAL and OUTPUT APPLIED on the other. In Model 2, however, the independent variables show up the distinction between OUTPUT APPLIED on the one hand and the THEORETICAL and EXPERIMENTAL outputs on the other. The high mathematical bias in OUTPUT APPLIED is reflected in the high path coefficient from BURSARY MATHEMATICS.

As the correlation between total scores (final examination plus course work) was very high (0.90 in 1973, 0.86 in 1974) consideration was given to the prediction of the grand total physics mark (total mark from the two courses combined ~ equal weighting). In Table 11.3 and Table 11.5 the regression factor structure coefficients of this PHYSICS GRAND TOTAL with the four objective input variables (Model 1) and the two "subject" input variables (Model 2) are considered respectively.

11.2 Content model

Although both the seventh form physics course and the first year university physics courses stressed the unity of physics and broke down the old established boundaries of heat, light, sound, and so on, three broad but distinct content areas are identifiable in these courses. These three areas are mechanics, fields and waves.[5]

The input and output tests were designed so that equal emphasis was given to each of these content areas. This allowed for any differences in attainment of the students over these three areas of content to be investigated by the use of the same type of questions

---

[5] In the seventh form course these three aspects of content are emphasized in the teacher's guide appendix (see CDU, 1973, p.108). In the first year theoretical course the lecture programme is split into six lecture-units, two on each content area. In the first year experimental course there were three groups of four experiments. Each group was primarily in one of these three areas (see Osborne, 1974a).

covering the same course objectives.

The objectives dependent on content were primarily Objectives I, II and III and the questions on each objective were designed to cover the three content areas equally. Hence (see Appendix K.1) the 36 multiple-choice questions covering Objectives I, II and III on the input tests contained twelve multiple-choice questions from each content area with approximately the same number of questions on each objective. It was therefore possible to obtain a MECHANICS, a FIELDS and a WAVES total which were different in terms of the content covered but involved the same question-type (multiple-choice) and involved the same objectives. Similarly for the output tests, the 45 multiple-choice questions covering Objectives I, II and III, provided fifteen questions in each content area (see Appendix L.1) covering the same objectives with the same question-type.

The coefficients of reliability of the content totals were estimated using the split half method and the Spearman-Brown formula (Guilford, 1965, p.457) to be between 0.60 and 0.70 for the input scores and between 0.70 and 0.75 for the output scores.

The three-input to three-output path diagram was postulated to have all possible paths. As only one path was not statistically significant at the 5% level, all paths were retained and the one not significant path indicated by an asterisk. Figure 11.3 shows the path model, Model 3. Tables 11.6 and 11.7 provide the associated correlation matrix and regression factor structure coefficients respectively using the conventions adopted in Section 11.2. As all paths have been retained the only differences between observed correlations and correlations associated with the model occur because the independent variables do not account for all the correlations between the output variables. The residuals are correlated.

Model 3 does not show significantly larger path coefficients between like input and output content areas than between unlike input and output content areas (at least at the level of reliability available with these 12 and 15 multiple-choice scores). One might propose that basically all path coefficients are likely to be of similar strength, particularly when the factor structure coefficients are considered and that other second order influences account for much of the observed differences in path coefficients. For example, because of the increased difficulty for students when questions in different content areas are in a random order, questions in both input and output tests were placed in the order mechanics, fields and waves. The high value path coefficient from INPUT WAVES to OUTPUT WAVES may well have been inflated as both the WAVES scores may have included a speed or overall "test competence" factor.

It was possible to investigate this content model further as the details of the success of individual students on individual questions in the Bursary physics examination of the preceding year were available. Totals were obtained for BURSARY MECHANICS, BURSARY FIELDS and BURSARY WAVES for each candidate. As there were insufficient multiple-choice questions in some content areas to provide reliable totals each content total included some free format answers. Questions not clearly on a particular content area, or of doubtful validity according to the examiner's reports, were not included. A score out of approximately 25 marks was obtained for each of the areas—mechanics, fields and waves—each with a coefficient of reliability of at least 0.70.

Again all possible paths were included in the path diagram. Only two paths were found to be not significant at the 5% level. These were not deleted but they are shown in the final path model with asterisks (Figure 11.4; Model 4). As all paths were retained the correlation matrix of Table 11.8 is symmetric apart from the effects of residual correlations. All regression factor structure coefficients were large (Table 11.9). Again, significantly large path coefficients do not consistently link similar content areas.

In comparing Model 3 with Model 4 certain patterns are clear. For OUTPUT MECHANICS, INPUT MECHANICS provides the largest path coefficient in both Model 3 and Model 4. Such a pattern might be anticipated for all content areas. However, INPUT FIELDS in fact provides the smallest path coefficient in both models for OUTPUT FIELDS. The reason for this can only be conjecture at this stage. However, teachers have often commented on the difficulties they have in teaching "fields" and the INPUT FIELDS score may not be representative of the student's ability because of the diversity of school instruction in this content area. Moreover, the end-of-test effect proposed to account for the high path coefficients from INPUT WAVES to OUTPUT WAVES is confirmed in that the Bursary physics papers do not follow the sequence of the tests used for Model 3 and a large number of FIELDS questions appear towards the end of the Bursary examination papers. Hence on the basis of this reasoning the path coefficient between BURSARY FIELDS and OUTPUT WAVES might have been anticipated.

Tables 11.7 and 11.9 show the regression factor structure coefficients and the multiple correlation coefficients,with PHYSICS GRAND TOTAL as the dependent variable and the input content scores as the independent variables. These results suggest that with these students in these courses a deeper knowledge of student achievement in any one content area is unlikely to provide a better predictor of future success in the course as a whole than an equally reliable measure with content coverage from the entire course.

11.3 A question-type model

The output examination for the theoretical course involved three types of examination questions;

(i) 30 multiple-choice questions (2 min each)
(ii) 9 short questions—free-format-answers (6 min each)
(iii) 3 long questions—free-format answers on a particular theme (20 min each).

These three groups of questions, each covered the three content areas, and attempted to assess the same objectives (Objectives I to III, see Appendix L.1). In the input tests only multiple-choice questions were used to assess these objectives. In the Bursary physics examinations two types of questions, "multiple-choice" and "free format" were used. Each of these groups of Bursary questions covered the three content areas and emphasized objectives similar to those of the final theoretical examination. This allowed some investigation of a question-type model.

The coefficients of reliability of the Bursary free-format totals and multiple-choice totals were each estimated using the Spearman-Brown formula and found to have a lower limit of 0.85 and 0.80 respectively. The coefficients of reliability of the final theoretical examination subtotals (long question, Section A; short question, Section B; and multiple-choice, Section C) were similarly assessed to have a lower limit of 0.80.

All possible paths were included in the path model. Only one path was found to be not significant at the 5% level and hence the path was retained in the model but it is denoted by an asterisk (Figure 11.5, Model 5). Hence again the correlation matrix (Table 11.10) is symmetric apart from the effects of residual correlations. As can be seen from the path coefficients and particularly from the factor regression coefficients (Table 11.11) there was little significant difference in paths from the two inputs to any of the three outputs. The Bursary free format total which accounts for 66% of the Bursary paper can be expected to have a higher reliability than the Bursary multiple-choice total and hence to have a higher path coefficient to each of the outputs. The correlations of the Bursary free format answers (which included long questions and shorter components) with the three outputs 0.67, 0.68, 0.67 shows absolutely no differences between outputs in terms of this input. Model 5 suggests that for questions covering these content areas and these objectives, the multiple-choice score on the Bursary paper does not add any new dimension to the assessment of student achievement not covered by the free format answers. Further, when the time spent in the Bursary examination on the different sections (free format versus multiple-choice ~ 2 hr versus 1 hr) is taken into account it would appear that the knowledge of a student's output score on questions of different types is unlikely to be useful even for predicting the student's scores on similar question types one year later.

The 32 Likert items used to measure Objectives V(C), VI(A), VI(B) and VI(C) provided a third type of question for an input total for Model 5. However, these "questions" measured different objectives and hence could not be included in this model. For interest, however, the analysis was done and as expected this Likert item total did not provide any statistically significant path to any of the three outputs of Model 5.

In summary, three distinct models have been considered in this chapter. The objectives model was investigated using both the input test scores, and the Bursary results, as inputs. The similarities and differences between Model 1 and Model 2 show how a different view of the system, obtained by using different input variables, may enable a better and broader perspective of the system to be obtained. The content model suggested how second order influences such as the position of questions on the examination paper may influence observed relationships. Both the content model and the question-type model indicated that, in general, measured student attainment is unlikely to be strongly dependent on either the content, or the question-type, used to measure this attainment.

CHAPTER 12

BACKGROUND VARIABLES AND MULTISTAGE MODELS

OUTLINE: The potential of path analysis for longitudinal studies of educational systems is investigated. Multistage models of the first year instructional system are developed. Consideration is then given to how these initial, necessarily crude, models could be improved in further studies. Finally, some of the significant features of the first year physics instructional models of Chapters 11 and 12, are summarized.

12.1 Background variables

In Table 10.2 a number of background variables were identified which may directly or indirectly influence output achievement in the first year courses. Of these variables the spatial aptitude and abstract reasoning scores were relatively highly correlated with each other in comparison with the correlation of either of these variables with any other variable considered (e.g., 0.62 compared with the highest other correlation of 0.34 in 1973). Hence to simplify path analysis these two scores were combined (equal weighting) to provide a single aptitude score for each student. In the following analyses this variable will be labelled APTITUDE (or APT). The lower limit estimate of the reliability of this measure is given in Table 12.1, along with the reliabilities of all variables considered for the first time in this chapter.

Similarly teacher's assessment of a student's "academic industry" and "maturity/stability" were relatively highly correlated compared with their correlations with other background variables. Hence to increase reliability and to simplify path analyses these two assessments were combined (equal weighting). In the following analyses this composite variable will be labelled MOTIVATION (MOTIV) on the assumption that an industrious seventh former with a high maturity/stability rating is likely to have high motivation for academic work (moreover, likely to continue to be motivated at university compared with a less stable individual). This is undoubtedly a crude indicator, but it was a direct and simple way of obtaining an indicator of a variable likely to be an important factor in academic achievement and one which was also considered important by students (see Section 10.1).

Two important background variables which affect achievement in the first year physics courses are undoubtedly achievement in physics and mathematics just prior to matriculation (BURSARY PHYSICS and MATHEMATICS). A number of other background variables identified in Table 10.2 can be postulated as affecting these achievements. Hence under the assumptions of Section 10.3, the multistage path analysis diagram of Figure 12.1(a) was postulated. The path coefficients were determined by multiple regression analysis using the mean correlations from the 1973 to 1974 data (N = 159). All paths not significant at the 5% level were deleted and the path coefficients recomputed. Figure 12.1(b) details the final path model.

Of the 23 postulated paths only nine were significant. From the path coefficients of Model 6 ~ Figure 12.1(b) the correlation coefficients given in the upper diagonal of Table 12.2 were determined. Again the "single and double underline" convention introduced in Chapter 11 has been adopted. Inadequacies in the model arise because of the fact that no causal relationships have been postulated between the time co-incident data. The correlation between School Certificate English (S.C. ENGLISH) and School Certificate mathematics (S.C. MATHS) cannot be accounted for by the independent variables, and there is a significantly high correlation of the residuals of S.C. ENGLISH with other dependent variables. In this particular study this is not too unacceptable in that S.C. ENGLISH could have been removed entirely from the model without major disturbance. With regard to physics achievement the removing of the S.C. ENGLISH would not have resulted in a significant loss of information. The other statistically significant correlation of residuals is between BURSARY PHYSICS and BURSARY MATHEMATICS. This shows that the preceding variables do not adequately account for the correlations between these two BURSARY totals.

While Model 6 has inadequacies it is nevertheless a useful, even if not unique, interpretation of the relationship between background variables. The significant paths are intuitively reasonable. For example, the indirect effect of APTITUDE on BURSARY PHYSICS ($0.34[0.47 + 0.30 \times 0.39] = 0.20$) via S.C. MATHS is greater than the direct effect (0.14). PARENT's education has no significant effect on any dependent variable. Females are slightly better at S.C. ENGLISH than males but otherwise there is no significant distinction between male and female with regard to any other variable considered.

If the model were to be developed further two possible procedures could be followed:

(i) Causal relationships could be postulated between time co-incident variables, and models based on different direction postulates could be tested in terms of the observed data and compared.[1] Although in this way it might be possible to account more adequately for the observed correlations, it may not add much to the understanding of how the system functions.

(ii) A search could be made for other independent variables

---

[1] For example, it could be argued that co-incident mathematics achievement is likely to affect physics achievement but that the converse does not hold.

which might more adequately account for correlations between the dependent variables.[2]

At the present crude stage of Model 6 the second procedure would be a more useful way to improve the model. What is shown by Model 6 is the potential of path analysis for longitudinal studies in that it provides a technique for evaluating the adequacy of the model in terms of how it accounts for the actual correlations between variables. From this crude model it is clear that early identification of important independent variables, and the measurement of these variables at the correct point in time, are of prime importance.

12.2 Components of the Physics Grand Total

The correlation between the total grade scores in the two physics courses was unexpectedly high (0.90 in 1973 and 0.86 in 1974).[3] However, although the assessment of the EXPERIMENTAL objectives occurred only in the experimental course, the assessment of both the THEORETICAL and APPLIED objectives occurred in both courses and there were other similarities between the courses. For example, both courses covered similar content areas; the course work in both courses was assessed by four hours of tests during the year; the format of the final examinations and the way the total grade score was computed were identical for both courses. The high correlation suggests that the basically different styles of teaching in the two courses did not greatly affect the total grade scores.

Because of the high correlation between the total grades in each course, a single score, the PHYSICS GRAND TOTAL (the simple sum of the

---

[2] For example, a measure of students' motivation for academic work prior to School Certificate might be postulated as a more useful independent variable than PARENT's education.

[3] In both courses the total grade score was the total of the final paper plus course work in the ratio of 2:1.

final total in each course), was considered in Chapter 11. This total was also of practical importance in that, in establishing the final grades of a student in the two physics courses, some compensation in grades was given to a student failing in one course who had a high mark in the other.[4]

Duncan (1966) suggested that many variables in the social sciences can be regarded as composite variables, and where such a composite variable occurs it may be of interest;

(i) to compute the relative contributions of the components to the variation in the composite variable, and

(ii) to ascertain how causes affecting the composite variable are transmitted via their respective components.

In terms of Model 1 and Model 2 (Section 11.1) PHYSICS GRAND TOTAL can be considered to be made up of three components: OUTPUT THEORETICAL, OUTPUT APPLIED and OUTPUT EXPERIMENTAL. However, PHYSICS GRAND TOTAL was not a simple direct combination of the "three" component scores, in that scores in test questions and some examination questions contributed to PHYSICS GRAND TOTAL but were not included in the component scores. Nevertheless, all questions attempted to measure objectives included in the component scores and no others, except for 2.5% of the PHYSICS GRAND TOTAL which was an effort mark for the student's laboratory account. Hence the three OUTPUTS of Model 1 and Model 2 were not the complete components of PHYSICS GRAND TOTAL. Nevertheless, an analysis of the relationship between these three component scores and PHYSICS GRAND TOTAL was considered to be of interest.

The component model for the PHYSICS GRAND TOTAL was computed and

---

[4] The total grade scores considered in this research are not compensated grade totals.

all paths were statistically significant (Figure 12.2). The model was again based on the averaged correlation matrix for 1973 and 1974 (N = 165) (Table 12.3). The multiple correlation of 0.96 (see Table 12.4) seemed appropriate with only 8% of the variance not accounted for. The inclusion of the OUTPUT ATTITUDINAL score as a fourth independent variable in Model 7 did not produce a significant path coefficient and hence did not reduce the unexplained variance further.

The differences in path coefficients in Model 7 are accentuated by the relatively high correlation between the components. These differences can be compared with the similarity in the regression factor structure coefficients for the three components (see Table 12.4). However, the relatively high weighting on the OUTPUT EXPERIMENTAL was unexpected but was consistently present when the 1973 and 1974 results were analyzed separately.

The most likely reason for the high weighting on the OUTPUT EXPERIMENTAL is that a systematic bias has been introduced by the fact that 66% of the PHYSICS GRAND TOTAL came from free format answers. This question type was used to assess the EXPERIMENTAL objectives but not the THEORETICAL or APPLIED objectives. A path diagram using the three components with the final examination totals in the theoretical course (73.101) and the experimental course (73.102) separately showed that the OUTPUT EXPERIMENTAL path coefficient was much less for the 73.101 output than for the 73.102 output. But it was not negligible. As the theoretical course (73.101) examination contained no questions on the EXPERIMENTAL objectives these results suggest that an ability to answer free format questions has biassed the results towards the OUTPUT EXPERIMENTAL score.

12.3 Other multistage models

In Section 11.1 the learning system was viewed as an input-output system, in Section 12.1 a multistage view of the background variables was developed, and in Section 12.2 the components of PHYSICS GRAND TOTAL were investigated. Because of the increasing discrepancies from observed data which result from a complex and many-stage model, no attempt was made to combine aspects of say, Model 2, Model 6 and Model 7 into a complete single model. However, the three models when placed in juxtaposition provide an overview. This is shown in Figure 12.3. Other models using other input and output variables (for example, using the variables considered in Models 1, 3 and 4) could be used to develop similar overviews.

As the shading in Figure 12.3 shows, no within-the-instructional-system variables have been included in these earlier models. However, this could be done and, in fact, it might provide more useful information for decision-making purposes. As a tentative investigation of this possibility additional variables were included in a model (Model 8) along with the variables of Figure 12.3. The additional variables were:

(i) INPUT AIM (INAIM)—this was a student's stated grade aim at the beginning of the year. Each student was asked to state his/her aim at the beginning of the year for each course (see Appendix K.5). Again later in the year students were asked what their grade aim had been at the beginning of the year in each course. All four scores were highly correlated (typically greater than 0.8) and a simple combination of these (equal weighting) provided a single score.

(ii) INPUT ATTITUDE (INATT)—this was INPUT ATTITUDINAL used in Model 1 and Model 2. It was included as the earlier models indicated that this was a measure of a variable quite distinct from the six input variables of Figure 12.3 (Model 6).

(iii) INDUSTRY (IND)—lecturers subjectively assessed the academic industry and maturity/stability of the students in a way identical to that of the teachers (see Section 13.2). These two lecturer-assessments were highly correlated (~ 0.80) and the sum of these assessments (equal weighting) correlated highly with the assessment of the student's laboratory book (~ 0.75). A student's laboratory book was marked for completeness and hence was considered a good indicator of student industry at least in the experimental course. The sum of the "industry" and "maturity/stability" assessments was added to the laboratory book mark (equal weighting) to obtain an INDUSTRY assessment composite.[5]

Information was also obtained from students on their "term environment" and also on their "vacational environment" (see Questionnaires—Appendix L and Appendix M). No useful way was found to include this non-parametric data into any model. For example, no assignment of numerical values to the alternative environments (home, private board, and so on) could be found which simultaneously;

(i) yielded an even approximately normal distribution,

(ii) could be interpreted in terms of any underlying variables (e.g., security), and

(iii) indicated a significant correlation with any output variable.

For these reasons the variables were not included.

Rather than introduce the inadequacies of Model 6 into a more extensive model, no causal relationships between these input variables

---

[5] Subjective lecturer assessments of industry and maturity/stability were made on two occasions during the year. Laboratory assessments were made on three occasions. Average values were used.

was proposed. However, Model 6 may be useful in interpreting relationships found in the more extensive model. A path diagram, Figure 12.4, including all possible paths between the variables was postulated, with the exception of paths between time co-incident variables (i.e., between INAIM and INATT, and between the four final outputs). Multiple regression analyses were carried out for all dependent variables. Path coefficients not significant at the 5% level were eliminated and the regressions re-analyzed. This procedure was repeated where necessary and particular care was taken to see that marginally significant paths were not eliminated prematurely. The final path model (Model 8, Figure 12.4) retained only 15 of the original 56 possible paths. Table 12.5 provides the correlation coefficients. The model was unable to account for five observed correlations and these can be accounted for by significant correlations between residuals. These five significant residual correlations are shown in Figure 12.4.

In Model 8 (Figure 12.4) the output achievements are seen to be related to different background variables. BURSARY MATHEMATICS provides the greatest direct path coefficient to OUTPUT APPLIED, while BURSARY PHYSICS provides the greatest direct path coefficient to OUTPUT THEORETICAL and OUTPUT EXPERIMENTAL. INPUT ATTITUDINAL provides the only significant path coefficient to OUTPUT ATTITUDINAL. The negative path coefficient from SEX to INPUT AIM (INAIM) suggests that female students have a lower stated self-perception of their ability than males. The not significant paths from SEX to all other dependent variables in Model 8 (Figure 12.4) and in Model 6 (Figure 12.1)[6] indicates that this perception is not justified.

MOTIVATION as assessed by teachers is shown in Model 8 to be

---

[6] Apart from S.C. ENGLISH which was small but positive.

related to INDUSTRY as assessed by lecturers, and whether at Bursary level (Model 6) or first year university level (Model 8) these measures did provide significant path coefficients to aspects of physics achievement in addition to the paths from earlier achievement.[7] APTITUDE as measured by the adapted DAT instruments appeared to provide no really useful information. The low but significant path from APTITUDE to OUTPUT EXPERIMENTAL in Model 8 (Figure 12.4) is consistent with the low but significant path from APTITUDE to BURSARY PHYSICS in Model 6 (Figure 12.1).

The combination of positive and negative path coefficients from BURSARY PHYSICS and S.C. MATHEMATICS respectively to INPUT ATTITUDINAL (INATT) is of interest and is consistent with a similar pattern of doubtful statistical significance between BURSARY PHYSICS, BURSARY MATHEMATICS, and OUTPUT ATTITUDINAL in Model 2 (Figure 11.2). The ATTITUDINAL scores in Models 2 and 8 did not include the "attitude to mathematics in physics", Objective VI(A), scores so the negative path coefficient from MATHEMATICS to ATTITUDINAL is not inconsistent. One interpretation of the relationship would be that if two students have comparable attainments in physics then the student who is mathematically weaker is likely to have a slightly more positive attitude to physics. However, it must be kept in mind that the PHYSICS and MATHEMATICS scores account for a very small amount of the variance in the ATTITUDINAL scores.

Finally, a student's stated grade aim at the beginning of the year (INAIM) provided a significant path only to OUTPUT APPLIED.[8] But even

---

[7] The relative magnitudes of the INDUSTRY paths to the different OUTPUTS are not inconsistent with what might be anticipated from a knowledge of the instructional system.

[8] Possibly a student's confidence is a common factor in both the stated aim and the ability to solve problems.

this was not particularly strong and overall INAIM made a negligible contribution to the model.

12.4 Input-output and multistage models of the first year physics courses

In Chapters 11 and 12, student attainment in physics has been considered from a number of different points of view (Models 1 to 8). The summary statistics provided by these models undoubtedly hide detailed information about individual students which may be of value, particularly as feedback to those individuals. However, such detailed information is likely to be most useful when considered in relation to overall system models.

Some of the main points brought out by the models (Models 1 to 8) include:

(i) ATTITUDINAL attainment as measured was quite distinct from cognitive achievement. Also measured OUTPUT ATTITUDINAL was strongly determined by INPUT ATTITUDINAL. The internal consistency of the attitude measures were adequate (Section 11.1) and considerable effort (Chapter 8) was made to develop a model with factorial validity (see Guilford, 1965, p.470). Indications, both in pre-testing items and in the 1973 and 1974 data, were that students who clearly demonstrated a very positive attitude to physics in their classwork, obtained high ATTITUDINAL scores. However, conflicting evidence on changes in attitudes (Section 9.2) and the isolation of input and output attitudes from other variables (Models 1, 2, 7 and 8) suggest that the measure itself may be inadequate.

(ii) With respect to cognitive achievement, Model 3 and Model 4 did not clearly establish stronger input-output paths between like content areas than between different content areas.

Model 5 did not suggest any stronger links between input-output paths on the same type of question. However, Model 7 did suggest that question type introduced some systematic bias into the comparison of the three cognitive outputs; THEORETICAL, APPLIED and EXPERIMENTAL. It is clear that once it is decided that a particular method is the best way to assess the achievement of a particular objective then the method may need to be stated along with a profile score to be most useful.[9]

(iii) As assessed in this study OUTPUT THEORETICAL and OUTPUT EXPERIMENTAL were strongly determined by earlier achievement in physics, as Models 1, 2 and 8 all indicate. OUTPUT APPLIED is more strongly dependent on MATHEMATICAL ability (Models 2 and 8). These patterns and the relationship of INDUSTRY to the OUTPUT scores in Model 8, raise the question, "What should be the desirable relationship between INPUT and OUTPUT achievements?" Very high path coefficients between INPUTS and OUTPUTS suggest that a course provides little opportunity for students to develop in different ways in that their success is predetermined by past success. On the other hand, very low path coefficients between INPUTS and OUTPUTS suggest that input measures are poor predictors of output achievement. The relationship of INDUSTRY to particular aspects of OUTPUT achievement is related to this question and is of importance.

(iv) A number of different models of a system are desirable. The

---

[9] For example; Score X is the student's ability to solve simple problems as assessed by multiple-choice questions, Score Y is the student's ability to plan experiments as assessed by short examination questions, Score Z is the student's ability to communicate as assessed from written reports.

distinctions between the OUTPUTS as shown by the relationships between INPUTS and OUTPUTS in Model 1 (Figure 11.1), for example, are different but not inconsistent with the distinctions between the OUTPUTS as shown by the relationships between INPUTS and OUTPUTS in Model 2 (Figure 11.2). Also extensive multistage models in education are not feasible because of the compounding of errors and the problems of measurement. More limited models in series may be more appropriate, e.g., Figure 12.3.

Finally, in interpreting models it would appear desirable that the "causal" interpretation of high path coefficients should not be stressed. A high correlation between two variables cannot be identified as cause and effect. However, where patterns of significant path coefficients reinforce other evidence (or even intuitive feelings) about a system, then such models may provide useful feedback to improve the instructional system, or suggest more detailed investigations of a particular aspect of the system. The relationship between MATHEMATICS and OUTPUT APPLIED, for example, would appear well worth further consideration and investigation.

CHAPTER 13

SUBSIDIARY MODELS

OUTLINE: Three subsidiary aspects of the first year instructional system are studied. A feedback model to take into account the changing aims of students during the year is investigated. The subjective assessment of student attainment in physics, by seventh form teachers and university lecturers, is analyzed using principal components and a simple diagrammatic model developed. The characteristics of students who do not complete their courses, and the future intentions of those that do, are briefly considered. Finally, some of the problems, and advantages, of developing specific statistical models are summarized.

13.1 A feedback model

A feedback cycle of possible importance which is suggested by the hypothetical generalized curriculum model of Chapter 4 is the interaction between the attainment aims of students and their test results. While it is true that in Model 8 a student's aim at the beginning of the course did not strongly effect output achievement, it is possible that this aim changed during the year and that a feedback model would more adequately represent reality.

In October 1972, 81 students were asked what their aim, in terms of grades in first year physics, was; (i) at the beginning of the year, (ii) at mid-year, and (iii) at the end of the year. As shown in Table 13.1, 38% of student replies showed a significant change of aim during the year.[1] The main reason given for a change in aim by the surveyed students was "physics tests and results from tests". Of the 81 students, 50% considered their test marks were lower than

---

[1] In addition the 10% of students who dropped the course had presumably had a considerable change in aim.

expected and Table 13.2 shows that there was possibly some relationship between actual versus expected test results and a student's change in aim.

Although the 1972 survey did not show that the test result-aim feedback loop was particularly strong, it did indicate that some feedback did occur and this aspect of the generalized curriculum model is isolated schematically in Figure 13.1(a). With respect to path analysis, Wright (1960b) has considered some aspects of feedback, in particular standard feedback parameters. In the context of educational systems with limited data and inaccurate measures such an approach appears to be inappropriate, as Duncan (1966) has suggested.

An alternative approach would be to consider each situation individually and to respecify Figure 13.1(a) in a manner more appropriate to a path analysis treatment. For example, if it is assumed that individual test results and "aim" interact, then in the theoretical course where there are four tests during the year and a final examination the path diagram of Figure 13.1(b) could be proposed. To investigate this idealized model fully, suitable data measuring a student's aim in the course would be required. However, as an exploratory investigation, path analysis was applied to an even simpler model for which data was available (see Figure 13.1(c)) for both the theoretical course (73.101) and the experimental course (73.102). In addition to the BURSARY PHYSICS (BPHS) results, the total test marks were available for each student in each course (TESTS), as well as each student's stated aim in each course at the commencement of the course (AIM) and at the end of the course just prior to the final examination (OUAIM). The student's total mark in the final examination in each course was also known (PAP).

Based on the mean correlations for 1973 and 1974 (N = 165), a multiple regression analyses yielded the path coefficients of

Figure 13.2(a) and Figure 13.2(b) for the theoretical course and the experimental course respectively. The results show that past achievement affects a student's aim to a much greater degree than student's aim affects achievement in both courses.

The path model is "over identified" (see Appendix B), hence the model can be tested by comparing actual correlations with those predicted from the model (Table 13.3). The discrepancies between the double underlined correlations from the model and the actual observed correlations in the lower diagonal show the inadequacies in the model, in that it cannot account in particular for the correlation between BPHS and PAP in either course. Further analysis shows that the inadequacy of the model results from the omission of a direct link from BURSARY PHYSICS to the final paper performance (PAP). Inclusion of this link in the path diagram and recomputation of the path coefficients (Figure 13.3) result in correlations from the model which are much closer to the observed correlations (Table 13.4). It might be postulated that a component of BURSARY PHYSICS and final paper performance (PAP) is simply ability in a final end-of-year examination.

13.2 Teachers' view of student attainment

Teachers have a store of knowledge about pupils which is developed from observing a wide variety of student responses, as Harlen (1975) and others have noted. The response of a student in class discussion, in lectures, and in the laboratory, and his/her responses to problems, self-evaluation exercises, tests and assignments all affect the teacher's view of the student. If this information is to be useful for feedback or prediction it needs to be quantified. Ideally, any subjective assessment of students by teachers should be done at the time of teaching. However, even in retrospect, a teacher's subjective assessment of a former pupil might well provide useful input variables for a model of the learning system, particularly in areas such as motivation where alternative measures are difficult to obtain.

As mentioned in Chapter 10, teachers were asked to assess retrospectively and subjectively their previous students' academic industry and maturity/stability in the seventh form (see Appendix N). At the same time, teachers were also asked to subjectively assess their former seventh form pupils on other variables of potential interest. In addition to industry (labelled U2) and maturity/stability (U3), were academic ability (U1) and the major complex of objectives emphasized in the first year courses (Table 5.1) which were similar to the seventh form goals. These were: knowledge and comprehension (P1), application (P2), critical analysis (P3), experimental work (P4), process experimental data (P5) and attitude to physics (P6). The "objectives" were explained in simple terms (see Appendix N) and of necessity more related to the seventh form physics goals than to the first year physics objectives. Nevertheless, apart from the omission of the complex "Nature of Physics", which is not really specified at the seventh form level, the groups of objectives of Table 5.1 correspond to the explanatory statement written for teachers.

The response of the teachers to the request for this information was excellent and replies were obtained for all students (N = 101 in 1973, N = 87 in 1974) with a seventh form background. To obtain an indication of the teachers' view of student attainment, a principal components analysis of teachers' responses was carried out on an ICL 4130 (Chelsea College) for each year separately. In both cases two significant factors were obtained (unrotated eigen values ⩾ 1.0). The first unrotated factor accounted for 63% of the variance in 1973; 69% in 1974. The second factor accounted for 13% of the variance in 1973; 10% in 1974. (The third factor accounted for 6.3% of the variance in 1973; 7.1% in 1974.) A Varimax rotation of these two factors resulted in the factor loadings shown in Table 13.5. Factor 1 is clearly associated with a student's "academic ability" as perceived by the physics teacher and Factor II could be interpreted as closely related to the teacher's perception of the student's "motivation towards academic work". Table 13.5 shows how teachers view the relative contributions of these two factors with respect to different objective areas.[2]

At the end of 1973, prior to the final examinations and without sighting the teachers' replies, two lecturers and a tutor/laboratory supervisor independently assessed those students they knew. At least two assessments were obtained for each student. Where there was a difference on a particular scale for a student on the independent assessments, an average was taken. In general, there was no major disagreement between lecturers. A principal components analysis showed that two significant factors were present in lecturers' assessments. (The first three eigen values were $\lambda_1 = 6.5$, $\lambda_2 = 1.3$, $\lambda_3 = 0.4$.) The first unrotated factor accounted for 73% of the variance, the second factor 14% (the third only 4%). A Varimax rotation of the two significant factors yielded the factors for 1973 shown in Table 13.6. There was a definite similarity between the lecturers' perceptions and the teachers' perceptions (Table 13.5) except that Objective P5 was seen by lecturers as more dependent on "motivation" than by teachers. This is expected, as less explicit pressure is likely to be placed on a student to complete experimental reports and data analysis at university than at school.

In 1974 the author made a particular effort to know all the first year students and visited them in the laboratories. From this observation students were assessed subjectively at mid-year and at the

---

[2] The relative importance of these complexes of objectives as perceived by teachers is given in Figure 5.3.

end of the year, again without sighting the teachers' assessment. The mid-year and end-of-year assessments were averaged (where necessary). A similar pattern again emerged from the principal components analysis with two significant factors (eigen values $\lambda_1 = 6.7$, $\lambda_2 = 1.2$, $\lambda_3 = 0.4$). Again, the first two factors accounted for 87% of the variance.

The similarity of the teachers' perceptions (Table 13.5) with the lecturers' perceptions (Table 13.6) shows that attainment is viewed subjectively by teachers and lecturers in a similar way. The "two-factor viewpoint" can be related to the simplified version of Figure 4.2 given in Figure 13.4. In this five-block model, "learner response" could be response in class, problems solved, experiments done, self-evaluation item responses, or competence with respect to a particular objective. Different responses are assumed dependent on different contributions from the different "inputs", x, y and z. For example, variation in "student responses" in terms of assignments within a class might be heavily dependent on student motivation. Alternatively, for certain responses (e.g., certain aspects of tests) the variation between students may depend more on background ability.

The results obtained in the components analyses are consistent with Figure 13.4 in that teachers (lecturers) when asked to assess their students subjectively will place emphasis on student academic ability (input y) and student motivation (input z). In any assessment of their own students on a national basis, the teachers are unlikely to give adequate weight to their own effectiveness as teachers (input x). In fact, correlations between any subjective assessment and any indicator of teaching effectiveness such as seniority, position of responsibility, degree qualification, and highest qualification in physics, all yielded correlation coefficients which were not significant.

The components analyses (Tables 13.5 and 13.6) show that two factors "ability" and "motivation" for academic work account in different proportions for most of the perceived variation amongst students with respect to all physics objectives (approximately 80% of the variance).

In Figure 13.5 a path analysis was developed using the two-factor scores[3] from the subjective assessments of teachers as independent variables and the objective outputs of Model 1 as the outputs. The low path coefficents from "MOTIVATION" FACTOR to OUTPUT THEORETICAL support the findings of Model 8, where a low path coefficient from INDUSTRY to THEORETICAL was observed. The multiple correlation coefficients in Table 13.8 are consistently lower than those of Table 11.5 but otherwise they follow a similar pattern of variation between outputs.

In Figure 13.6 the mean correlations between the subjective student assessments and the students' total marks in the two physics examination papers at the end of the first year courses are shown. The teachers' assessment was used to predict performance at the end of the subsequent year, while the lecturers' was for the same year. However, Figure 13.6 clearly shows strong similarities in the variation in the correlation coefficients across objectives (teachers' correlation + 0.2 ≈ lecturers' correlation).

Finally, an analysis of a combination of examination results and subjective assessments showed that the inclusion of both subjective factors along with BURSARY PHYSICS in a linear regression analysis did not significantly increase the prediction of the OUTPUT variables

---

[3] Factor scores were obtained for each year separately, correlations computed for each year separately using these factor scores and then these correlations were averaged (Fischer's Z and weighted according to student numbers).

(see Figure 13.7 and Table 13.9). The "MOTIVATION" FACTOR in particular, made no significant contribution to prediction.

13.3 Other models and summary

Much educational research has dealt with the problem of dropouts from university. This problem varies considerably from country to country and is dependent on entry requirements for university and the related necessity or otherwise of "sieving" first year students. In New Zealand a reasonably "open door" university policy exists and some dropouts from university are expected within or after the first year.

With respect to the first year physics courses in 1973 and 1974, of the 188 students who had a background of seventh form physics and took both courses: 10 students dropped out of university completely before the final examination (3 in 1973, and 7 in 1974) and sat no papers in physics or any other subject; 4 students dropped both physics courses (2 of these students dropped courses in other subjects as well); 3 students dropped the theoretical physics course only (all these students dropped courses in other subjects); 5 students dropped the experimental course only (4 of these students dropped courses in other subjects). Hence 22 students in the two years failed to sit the final examination in one or both courses and were therefore dropped from the analyses of Chapters 11 and 12. Of these 22 students only 4 dropped just physics courses; hence it would appear that the reasons for failing to sit one or both final examinations rests, at least partly, on factors "outside" the physics learning system.

Despite this it was of interest to investigate tentatively whether or not any of the input variables used in the Models 1 to 11 related in any significant way to the inability of a student to complete both courses. A dichotomous variable, labelled COMPLETE, was generated with a value of 1 if a student completed both courses and a value of 0 if a student did not complete one or both courses. This skewed dichotomous variable must be treated with considerable caution but the relative importance of different input variables using COMPLETE as the dependent "output" variable was evaluated by developing three path models (Figure 13.8). As expected none adequately predicted COMPLETE, but the relative importance of INPUT ATTITUDINAL and MOTIVATION FACTOR shown in Figure 13.8(a) and in Figure 13.8(c) is of interest. In Figure 13.9 the simple correlations between the teachers' subjective assessment of physics objectives and COMPLETE show a markedly different profile from Figure 13.6 and again the importance of "ATTITUDE" (P6) is emphasized.

As part of the final course questionnaire, students were asked to indicate their intentions with regard to pursuing further physics or physics-related courses for the following year (see Appendix L.2, Part C, Q3). Numerical values 5 to 1 were ascribed to responses (a) to (e) respectively. This provided a crude measure of the amount of physics intended in subsequent study. This variable FUTURE PHYSICS was approximately normally distributed (mean = 3.0, standard deviation = 1.0) and three path models were developed (Figure 13.10). Again, no model "explained" more than 10% of the variance (Model 1(b)). However, the relative importance of INPUT ATTITUDINAL to the cognitive inputs in Model 1(b) is in distinct contrast to its relative importance in Model 1. Correlation coefficients of subjective assessment versus FUTURE PHYSICS show a flat low distribution (Figure 13.11) significantly different from Figure 13.6 and Figure 13.9.

Particularly with regard to some of the aspects of the instructional system investigated in this chapter it could be argued that alternative methods of investigation and statistical analyses might have been more pertinent. However, in PART C the intention has been to undertake investigations from a broad elucidatory perspective which could be related to the generalized curriculum models of

Chapter 4, rather than to be particularly concerned with details of a more evaluative or specific nature.

From a consideration of the models developed in PART C certain problems of developing statistical models have been clarified. First, one of the major problems is undoubtedly the difficulty of identifying important variables and obtaining reliable and valid measures of these variables. Secondly, even if pertinent variables are recognized it is not always possible to measure these adequately because of the likely perturbations to the system caused by such a measurement. Thirdly, any particular statistical model which is used, is likely to restrict the type of data which can be effectively used in the model. Finally, it would appear from this study that the effective feedback obtainable from such models to improve the instructional system is unlikely to be commensurate with the time and effort put into elucidatory studies as compared with evaluative studies.

On the other hand, there are advantages in the attempt to develop statistical models in that it focuses attention on viewing the system in broad perspective, and it leads to an appreciation of the need to develop valid and reliable measures, particularly of student achievement. Moreover, even crude models are likely to give some valuable insight into overall relationships between variables.

From a long-term view the continual monitoring, development and modification of a system model over a period of years would appear to be necessary for useful and effective model building.[4] However, in the developing and refining of models certain basic questions would

---

[4] It is proposed that initial exploratory investigation should lead to simple models confined to a few specific variables of major interest. The use of methods such as path analysis could then identify major inadequacies in the models. Further development would lead to more pertinent variables and models which more adequately represent reality.

arise about which decisions would have to be made. First, is the unexplained variation in dependent variables the result of the wrong set of independent variables being chosen or of an inadequate set of too few variables being specified, or simply of the inaccurate measurement of correctly chosen variables? A decision on this will subsequently affect how the model develops. Secondly, if there are changes in the relationships between variables in the system (the model) from one year to the next, is this due to changes in the instructional programme, to changes in the student population, or to changes in the measurement of the variables? A decision on this will determine how information from the model might be used. Thirdly, if stable and clear relationships between variables are identified by the model, what inferences can be made from general experience or further investigations into the likely reasons for these relationships? A decision about the causes of a relationship will determine what changes are made to objectives, teaching and learning methods, and assessment.

PART D

AN EVALUATION OF THE SYSTEMS APPROACH

CHAPTER 14

SOME SPECIFIC APPLICATIONS

OUTLINE: The applicability of some of the aspects of the systems approach to senior secondary school physics teaching, is investigated. Specific examples are cited of how ideas developed in this thesis have been, or could be, applied at the secondary school level.

14.1 Objectives and instructional design

The first year physics courses at the University of Waikato are part of a wider physics education system summarized in Figure 4.2. Of particular interest and relevance to first year university teaching is the national senior secondary school courses in physics and during the period 1970-1974 the author of this thesis was also actively involved in course design, developmental work, and assessment at the school level. Specific examples will be cited in this chapter as to how some of the ideas and perspectives developed from a study of the systems approach to first year university physics have been, or could be, applied at the senior secondary school level in physics.

In Section 5.2 a case was made for the specification of general course objectives, as a means of orientating and diversifying teaching procedures, organizing and diversifying testing, and also as a means of communicating with other teachers. As stated in Section 5.1, goals have been stated for the national physics syllabus but barely at an operational level for either teacher or examiner. However, attempts have been made in the Teacher's Guides (e.g., CDU, 1973, p.iv-v) to make them more operationally useful and the work done on objectives in Chapter 5 in particular influenced this work.

It was pointed out earlier that teachers tend to be task- or topic-oriented and lists of course objectives stated in the introduction to a national syllabus or teacher's guide are not likely to have much impact on teaching unless the relationship between the objectives and classroom activities, and between the objectives and the national examinations or assessment procedures, is clear. In an attempt to help teachers clarify objectives, relatively specific behavioural objectives are sometimes included in national teacher's guides at the beginning of each task or topic (for example, CDU, 1973, p.4). These can only be samples of some of the student behaviours which it is hoped that the unit of work will develop and this needs to be made clear. Further, if such task-related objectives are to be included, the relationship between them and the general learner objectives needs to be extremely obvious if teachers are not to be confused, or to reject all consideration of one, or both, sets of objectives.

The experience gained from using general course objectives with the first year university physics courses suggests that clearly defined course objectives that are largely content-free may provide the most effective framework within which teachers can develop their teaching programme. Objectives can then be related to teaching by including in the teacher's guide parenthetical notes alongside activity, or task-oriented details which give examples of the potential an activity has for developing particular objectives. In this way teachers would develop their own awareness of objectives, look for other potential opportunities within their task-oriented teaching for developing particular objectives, and possibly develop new activities to emphasize particular objectives.[1]

In terms of student assessment, examples of test questions attempting to discriminate on particular objectives can be produced. These show teachers how objectives can be related to the assessment. Nicholas (1974), working at the University of Waikato in 1973, developed a set of such questions for seventh form physics classroom use. It would be helpful if national examiners pointed out how their examination papers relate to course objectives. An example of this is found in the report on the Scholarships Physics Examination 1973 (see Appendix I.2).

In these ways teachers might become more aware of the national objectives of a course and the all-important relationships between objectives and both teaching and testing. Hopefully, this would lead teachers to a search for other unspecified opportunities to develop and assess particular objectives within their own teaching and testing programmes, and lead to the refinement and respecification of objectives at a level of specificity found to be personally appropriate.

The work of developing and refining educational objectives, and of designing instruction in terms of these objectives (Chapters 5 and 6) led to the application of these ideas to the development of a textbook for the final year of secondary school physics in New Zealand (Barber and Osborne, 1974). The four main groups of objectives of the first year physics courses (Table 5.1) corresponded to the four objectives of

---

[1] For example, the teacher's guide may state that pupils should derive the period of a simple pendulum. If one of the objectives of the course is to develop an awareness of limitations and approximations in theories, then a parenthetic note (pointing out that in this activity there is a good opportunity to develop this objective by emphasizing the assumptions made in the derivations and the limitations in the resulting equation) will help teachers relate objective to teaching. Simple annotations of the type shown in Appendix D.3 may be adequate. The above method of presenting teacher's guide material was proposed by the author to the revision writing team of the seventh form Teacher's Guide (Hogben House, 1973). Although agreed to in principle, in practice it did not eventuate.

the seventh form physics course (see CDU, 1973, p.iv). Experience gained from relating objectives to teaching led to the development of various distinct aspects in the textbook. Each aspect related primarily to one particular objective, and the overall plan (Table 14.1) was one which at least co-ordinated the activities of the authors involved in writing the textbook. It is believed that this resulted in a textbook which is in sympathy with the aims of the seventh form course and which will provide suitable support for a variety of classroom teaching styles.

14.2 Process and product evaluation

Eisner (1969b, p.131) considers that teachers often evaluate their own teaching by the extent to which "students appear engaged, immersed, caught up and interested in classroom activities". As Eisner states, it is likely that this emphasis on process evaluation is because teachers believe that engagement, intellectual and emotional immersion, is a better indicator of educational value than the achievement of test scores. Undoubtedly process evaluation is also considered valuable by teachers because it provides immediate feedback on instructional methods and clear indications of "desirable" changes.

Process evaluation in a national secondary school system within a subject, can be at two levels. First, at the classroom level a teacher evaluates his/her own classroom teaching and management. Have students found particular units of instruction enjoyable or unenjoyable, informative or uninformative? Much of this evaluation is subjective and informal and the success or otherwise of classroom teaching is likely to be dependent on a teacher's receptiveness to feedback via students' stated or even unstated reactions. Informal feedback can be supplemented with more quantitative monitoring or formal questionnaires (Chapter 7) and some teachers use these procedures. However, the usefulness of such information for decision-making is limited by the small number of students involved (N < 25, say) and free comment by students either written or oral is likely to be more useful than forced-response type questionnaires.

Secondly, at the national level, information from teachers on the suitability of aspects of the curriculum for classroom teaching can be useful in the formative stages of developing a curriculum. This is particularly so if the national curriculum is in a unit form where it is desirable to identify the "good" units for emulation and the "bad" units for modification or change.[2] Forced-response questionnaires or even semantic differential techniques (Chapter 7) may be useful.

In addition to evaluating engagement, it is desirable to evaluate the extent to which identified or chosen goals are being achieved. Product evaluation at the classroom level is even more difficult than at university, in that small numbers preclude the use of statistical techniques to develop questions and analyze results. Specific objectives with criteria-based tests might be used for certain aspects of the work but, as Ebel (1971) states, in addition to the questionable desirability of such objectives there is also the doubtful validity of the criteria used by individual teachers.

The potential of national examinations for providing useful product evaluation feedback to individual teachers has not yet been realized. In writing examiners' reports there have been attempts (for example, see Appendix I.2) not only to provide details of the objectives assessed by particular questions but also to provide overall attainment statistics (means, standard deviations and correlations). Such

---

[2] Such a comparative investigation was begun by NZEI on units of work in a new Primer 1 to Standard 4 syllabus. Replies from a teachers' questionnaire were analyzed at the University of Waikato in 1973 and a particularly "good" unit was identified for the Curriculum Development Unit.

statistical information provides a "criteria" level set by the average national attainment. What would then be useful to the classroom teacher in terms of product evaluation would be information on the performance of his/her pupils in terms of the national average for individual questions, content areas, or groups of questions related to particular objectives. For example, mean profiles for a class compared with the national profiles in different objective areas (e.g., THEORETICAL, PROBLEM-SOLVING, EXPERIMENTAL) could be useful to a teacher making decisions about where he should place more emphasis in his teaching.[3]

In England, Mathews (1974) has considered how product evaluation in the form of examination results might be used as an aid to a continuous process of curriculum development. Here the concern is to provide teachers not only with information as described above, but also with national standards. Details of examination results can give some indication of how appropriate the content, objectives and the examination questions themselves are in terms of the national system. However, discrepant results between expected outcomes of attainment and actual levels of performance are difficult to interpret in terms of whether or not the discrepancies are the result of an inappropriate curriculum, inadequate classroom instruction or inappropriate examination questions. If an examination is considered appropriate by teachers, students and subject specialists and if student attainment is determined to be "reasonable" (e.g., little scaling is required to establish suitable pass rates), then the overall national product might be considered satisfactory. If this situation does not exist and it consistently remains unsatisfactory over a period of years (as Mathews

---

[3] Further research of the type outlined in Chapters 8, 11, 12 and 13, as well as further development in the validity and reliability of test measures, is necessary before valid and useful profiles within a subject are likely to be available for the guidance of individual students.

states, one cannot tinker with national systems every year), then serious consideration needs to be given to whether or not changes need to be made to expectation levels or to the curriculum itself.

14.3 Specific systems models

The development by classroom teachers of statistical models of the type outlined in PART C are not realistic at the secondary school classroom level. This is not only because of the few students involved in any one instructional system, but also because the time and effort which must be put into developing such models is unlikely (in view of the work in this thesis) to be commensurate with the amount of useful feedback which a teacher might gain from such an exercise. Normally a teacher is fully occupied with developing instructional methods, with teaching, and with process evaluation.

At the national level, however, with a centralized data bank, path models could be developed. Such models could be of value for monitoring the educational system over a period of years and observing changes in the relationships between the components of attainment within a subject, between attainments in different subjects and between attainments and background variables or future attainments. The changes in system relationships could be compared with changes in curriculum or assessment methods. Drastic changes in relationships caused by changes in curriculum and assessment methods can otherwise occur largely undetected.

Analyses of national examination results (Appendix I.3) as well as the work described in Chapter 8 show that it is difficult to clearly distinguish, significantly different aspects of achievements within an examination taken by students from a variety of learning situations. However, it would appear that, in future, pupil assessment in schools will involve a variety of assessment techniques—for example, examinations, internally assessed practical work, project work, and so on.

Particularly if different techniques emphasize the measurement of different objectives, the relationships of background variables, and "future success", to these components of achievement could well provide useful information about the objectives, the teaching, and the assessment methods. In Appendix I.3 some of the possibilities and limitations of investigations along these lines within the present examination system are tentatively explored, using the results of the 1973 Scholarships Physics examination.

14.4 Generalized curriculum model and manpower

The model of Figure 4.2 clearly emphasizes the feedback nature of physics education in terms of manpower. Even if no attempt is made to develop any specific statistical model of this complex system, it is important at least to monitor the numbers in the various units of the system. Over the last twenty years there has been much discussion in New Zealand about the shortage of physics graduates, the consequential shortage of well-qualified physics teachers in schools, the consequential drop in standards, or interest in physics by school pupils, and the "downward cyclic process". In order to establish the shortage of physics teachers in clear statistical terms it was found necessary to develop a questionnaire for headmasters which would compare the number of physics graduates in schools with the number of mathematics and other science graduates. Details of this 1971 survey were reported in 1973 (Osborne, 1973).

The figures of Table 14.1 suggest that the relative numbers of graduates in different subjects in schools is likely to be comparable to the relative number of graduates produced in those subjects by the universities (unless some new factor is introduced, such as a differentiation in incentives, selection or training procedures between subjects). The recent trends in the numbers of physics students in schools (Figure 14.1), in the number of first year university physics students (down 30% from 1972 to 1974) and in the number of graduates produced (Figure 14.2) suggest that physics is likely to continue to have manpower problems in schools.[4] Recent fluctuations in the numbers of graduate physics students entering teacher training in Britain suggest that financial considerations and the supply of alternative jobs for physicists are two important factors affecting recruitment of teachers. While a continual monitoring of the manpower situation is desirable, it is becoming increasingly difficult to isolate the physics education system from other subjects. The phasing out of the old University of New Zealand unit system means that it is difficult to assess the numbers of "physics students" in universities, or even the numbers of physics graduates being produced.[5]

Figure 4.2 also emphasizes the desirability of monitoring the subsequent courses, or careers, of physics students at each level (e.g., final year secondary physics students). This information could well be helpful in decision-making about course content and the objectives of a particular national programme.

In this chapter consideration has been given to how some of the ideas and viewpoints about learning systems, developed from a systems approach to the development and analysis of a university physics course, can be applied in the development and investigation of a national secondary school physics course. Undoubtedly, at least in New Zealand, much more could be done than at present in terms of the careful development and analysis of national learning systems in secondary school science.

---

[4] Bondi (1975) recently expressed the widespread concern about the drop in numbers of university physics students in the Western World.

[5] For example, University of Auckland B.Sc. graduates need no longer major in a particular subject although they must have an equivalent number of credits at third year level.

CHAPTER 15

CONCLUDING REMARKS

OUTLINE: The work described in this thesis is reviewed and the contribution which a systems approach can make to the development of an instructional programme is analyzed.

15.1 The systems approach

A basic reconsideration of the "systems approach" (Chapter 2) has led to a broader view of it than is normally considered in education. From this broader view two distinct aspects of the systems approach have been identified—an applied and a basic aspect.

With regard to curriculum development and analysis, this interpretation of the systems approach has led to a clearer perception of the different aspects of curriculum work and to a better understanding of the relationships between these aspects. The applied systems approach is concerned with the detailed design and evaluative aspects of curriculum development while the basic systems approach is concerned with broad perspectives and the elucidation of a developed instructional system primarily in terms of models.

Four different aspects of the applied systems approach to curriculum development can be identified—defining objectives, developing instruction, monitoring instructional methods, and assessing the product. These four aspects can be considered to be respectively supported by four areas of educational evaluation: context, input, process and product. With regard to the basic systems approach, learning systems can be considered from a wide variety of viewpoints and a generalized curriculum model makes explicit the particular subjective viewpoint from which more specific and necessarily limited mathematical or statistical models can be developed.

Undoubtedly the identification of the above six aspects of curriculum development, course evaluation, and system elucidation as isolated and distinct entities is an idealization in that they are interrelated, interdependent, and the boundaries are not always clear. However, in this research an attempt has been made, by argument and example, to show the importance and place of each and all of these aspects in the development, and understanding, of an instructional system.

In curriculum work, investigations can range from "informal and subjective" to "formal and objective" and this is detailed, with examples, in Table 15.1. Much elucidation and evaluation of educational systems occur at the "informal-subjective" level. Again, in this research an attempt has been made to show the usefulness of supporting informal and subjective comment with more formal and objective investigations.

15.2 The generalized curriculum model

A systems approach tends to emphasize particular viewpoints. With regard to the development of a generalized curriculum model of an instructional system, the systems view emphasizes;

(i) "the system" both as a whole and as part of a larger system,
(ii) the objectives-teaching-assessment aspects of instruction,
(iii) the relationships between inputs and outputs.

This approach to the development of generalized curriculum models led to the diagrammatic models of Figure 4.1 and Figure 4.2. These generalized curriculum models were particularly developed for the first year physics courses at the University of Waikato, and have been used as a basis from which to view all parts of this study. This led to a focus on the importance of the precise determination of objectives and teaching methods, and the use of student assessment and feedback loops within the system. The nature of educational variables is such that it is unlikely that mathematical models will ever replace entirely the wider and more subjective views of generalized descriptive models. In this particular study the specific system models did little to support or contradict these generalized models.

15.3 Objectives and instructional design

In the first year courses the systems approach led to the formulation of explicit objectives. These objectives were not intended to cover all aspects of physics education because the first year courses were seen as part of a three-year degree programme in physics. Certain objectives (in particular, experimental techniques and communication skills) were deliberately de-emphasized or omitted.

The first tentative objectives in teaching and student assessment (Chapter 5) were later refined and the optimum level of specificity for maximum usefulness determined. While specific behavioural objectives may be useful for the achievement of some objectives in individualized instructional systems, a very early attempt to write such objectives for particular content areas and learning activities in physics was abandoned as impractical and too restricting with respect to the first year courses. The complete specification of a university course by such specific objectives is clearly inappropriate.

It was decided to make explicit general learner objectives for the first year courses unencumbered by details of exactly what to teach, how to teach and how to assess. During the development of the course programme certain distinctions between objectives became clear, and were found to be useful in the improving of teaching procedures. Similarly in writing questions to be used for student assessment, certain other distinctions between objectives became clear. Twenty-two general learner objectives were finally isolated for the first year courses (Table 5.1) and, in practice, these objectives were found to co-ordinate and diversify both teaching and assessment.

The order and grouping of objectives (Table 5.1) were developed from teaching and also from designing questions for tests and examinations. A statistical analysis of student attainment of these objectives, however, led to an alternative grouping being proposed (Table 9.7). The four statistically established groups have some empirical basis in terms of student attainment. Compared with the seven groups of Table 5.1, these four groups may well be more easily interpretable by teachers and students, and be more useful when different aspects of student attainment are being considered.

With respect to instructional design a variety of activities were initially timetabled in the first year physics courses to cover the original implicit objectives.[1] Subsequently objectives were explicitly identified and teaching methods were adapted and changed whenever opportunities were seen to better achieve particular objectives (Chapter 6). Changes were also made in the light of information from both formal and informal process and product evaluation (Chapters 7 to 9).

The learning activities originally adopted for the course provided sufficient diversity of activity to permit all subsequent changes to be made without major reorganization. One of the main problems of instructional design is the difficulty of obtaining sound input evaluation on the basis of which the teacher can determine what teaching method is likely to be the most appropriate, or effective, for the

---

[1] Large lecture groups (N > 100), medium sized laboratory groups (N = 30), small tutorial groups (N = 16), independent study in learning aids room, informal self and small group study.

purpose of achieving particular objectives.[2] This difficulty is associated with the difficulty of obtaining effective product evaluation of other teaching programmes. At present, it would appear advantageous in practice to adopt a diversity of teaching methods in new courses which are to have a wide variety of teaching objectives. Subsequent process and product evaluation can then provide information about how best to use these methods to achieve particular objectives.

An associated aspect of instructional design is the determination of student input abilities and attainments, so that instruction can be designed more effectively. Although the input tests in this research were not primarily designed for input evaluation, or to provide feedback to students about the limitations of their prior knowledge, the potential of input tests for these purposes later became clear. The limitations of criteria-based tests in terms of product evaluation have been described in Section 8.1. However, in terms of input evaluation they would appear to have potential for the determination of deficiencies in the required prerequisite knowledge of students. Once determined these deficiencies could then be overcome by individual remedial learning or, if common, by modification of the instruction in the course.

Also of significance in instructional design is the relationship of the course assessment procedures to the learning activities and the objectives. In the first year courses the initial design, which involved giving equal emphasis to lectures and laboratory work by assessing the learning based on these two activities separately, was particularly successful. Degrees consisting of small course units

---

[2] For example, is individualized instruction better than lectures for the purpose of achieving comprehension objectives? Is project work better than conventional laboratory work for the purpose of achieving experimental objectives?

provide a useful opportunity to relate separately assessed courses to learning activities rather than to content areas.

15.4 Process and product evaluation

In the first year physics courses many changes in the programme resulted from informal process evaluation. Informal comments from students, observation of students' reaction to instruction, test results and so on all contributed towards a complex subjective impression of how successful instruction was and an intuitive feeling for what changes might be required. This "folk wisdom of the classroom" was supplemented by more formal feedback via questionnaires.

Isolated numerical information obtained from formal process evaluation was not found to be particularly useful. However, comparative data of two kinds were useful. First, a comparison of students' reaction to one unit of worked compared with another was informative (e.g., different experiments or chapters in the textbook). These comparisons indicated possible changes which might lead to course improvement. Secondly, a comparison of students' reactions to aspects of a course over a period of time indicated the effect of changes (intentional and unintentional) in the system. This "monitoring" of a course can indicate clearly that certain changes are desirable (Chapter 7).

While a certain favourable reaction to a course is necessary before learning is likely to be effective, student reaction is not the only criterion of course success. First, students react differently to a course and only in a highly individualized system with maximum resources could one be expected to satisfy all students at all times. Secondly, the satisfaction of the student with the process must also be set against the satisfaction of the student, teachers, and society with the product. However, process evaluation is emphasized because it is relatively simple to undertake and can be closely related to the instruction. In this way useful feedback information can be obtained for decision-making about specific system changes.

On the other hand, product evaluation of a learning system appears to be much more difficult. Nevertheless, aspects of product evaluation associated with student assessment in the first year courses did provide feedback for changing the instructional methods and the assessment procedures themselves. The explicit formulation of course objectives was found both to orientate and diversify assessment and to increase the usefulness of the feedback from student assessment. Factor analyses of early tests and examinations (1970-1972) showed that marks gained by students on a particular physics question appeared to be a complex function of the objective being assessed, the content covered by the question, the style of question used and the position of the question in the test or examination paper. Hence, in order to get a measure of attainment of a particular objective, the most appropriate assessment technique for each objective had first of all to be decided upon. Then a number of questions was developed for each objective in order to eliminate as far as possible the effects of the content covered by the questions, and the position of these questions in the paper, on the total mark for each objective. Subsequent analyses (1973 and 1974) of the attainment of the students on these measures indicated that the physics attainments were multidimensional (Chapter 8). But it must be re-emphasized that these analyses were dependent to some extent on the assessment procedures used to measure each objective.

The problem of assessing the course effectiveness in terms of the product and hence the associated problem of ascertaining what teaching method is best to achieve certain objectives, is a major problem of education from a systems viewpoint.[3] A clearer conception of objectives and of the different aspects of achievement in learning systems may ultimately lead to greater success in research in this area. Even where no significant differences in overall achievement (grand total) may be observed in comparative studies between different teaching methods, clear common objectives and effective measures of attainment of the objectives may well lead to the observation of significant changes in attainment of some objectives. Until there is clarity of objectives, and effective measuring techniques are available, a greater emphasis on a more subjective evaluation of the product by teachers may be desirable. For example, at school level a questionnaire technique similar to that used in process evaluation asking teachers for a subjective comparison of the effectiveness of units of work to develop particular objectives in students, may be more useful for measuring course effectiveness than the results of directly assessing the students.

Although not followed-up in this study, it would appear desirable with regard to product evaluation techniques to evaluate the product in the long-term. For example, questionnaires sent to former students to determine which attainments or learning activities had subsequently been most beneficial, may well provide useful feedback for product evaluation. Two difficulties arise, however. They are: first, to obtain replies from a representative sample of students; and secondly, to assess whether or not changes made to the learning system since the student was involved in it have not invalidated the student's comments.

---

[3] For example, comparative studies of different teaching methods have yielded few clear results. Of the 30 research studies reviewed by Welch (1972) which compared instructional methods, 17 studies indicated no significant difference and 6 studies indicated mixed results. Welch (1972) was critical of the conclusions drawn from many of the other studies, where significant differences were reported.

This type of study is also closely related to context evaluation and the choice of course objectives.

15.5 A specific systems model

Finally, the attempt in this research to develop statistical models of the first year physics courses which would aid in the understanding of the teaching and learning system, elucidated a number of significant factors:

(i) There are major problems of identifying and reliably measuring pertinent variables.

(ii) The practical constraints on the choice and measurement of variables when dealing with systems involving people are likely to be significant.

(iii) There is likely to be a need for a number of different models to clarify different aspects of the system.

(iv) The choice of a particular statistical model is likely to limit the type of data which can be usefully used.

(v) Path analysis would appear to have potential in educational analyses.

(vi) Models are likely to emphasize the inadequacy of certain aspects of student assessment and product evaluation.

(vii) Initially models are likely to be inadequate to provide effective feedback for course improvement.

(viii) Even crude models require a major commitment in time and effort which is unlikely to be commensurate with their immediate usefulness.

However, in attempting to develop specific models of the first year physics courses certain valuable insights have been obtained with regard to;

(i) the complex relationships between the measured student attainment and the measurement technique used,

(ii) the distinctions, and the relationships, between different aspects of measured achievement,

(iii) the relationships, even if only tentative, between background variables, and different aspects of achievement.

In conclusion, the work done on the development of specific system models has led to strong agreement with the view put forward by Blalock (1964). This view is that specific model building in natural settings is likely to be, at present, of limited practical value because the measurement techniques, the control over extraneous variables, and the theoretical tools available may not yet be adequate to the task. However, if in some point in time we are to successfully analyze and understand non-experimental situations, we must continue to work in this area to identify, and face up to, the problems involved.

15.6 Conclusions

The present study has shown, by argument and example, how a broad systems approach can be applied to the development and evaluation of an instructional programme. Some of the possibilities, some of the problems, some of the limitations of this systems approach in a teacher developed and evaluated system have been identified.

Much more remains to be done to show the full potential and implications of a systems approach in education. However, in this study, a systems approach to physics education at the first year university level has;

(i) provided a broad and comprehensive framework on which to base curriculum work (Chapter 2),

(ii) provided a viewpoint which led to an emphasis on particularly important aspects of an instructional system; namely the objectives, the teaching methods, the assessment procedures, and the feedback mechanisms (Chapter 4),

(iii) ensured that the objectives of the course were clarified and explicitly stated (Chapter 5),

(iv) emphasized that teaching methods were considered in terms of their potential for the achievement of course objectives (Chapter 6),

(v) emphasized the importance of obtaining useful feedback information from students, so that appropriate changes could be made to instruction and so that the effect of these changes could be monitored (Chapter 7),

(vi) emphasized the importance of adequate measurement of the attainment of the course objectives (Chapters 8 and 9),

(vii) emphasized the need for a better understanding of the system through the development of specific statistical, or mathematical, models (Chapter 10) and,

(viii) provided guidelines for the development of specific models of the system (Chapters 11, 12 and 13).

In the case of the first year physics courses at Waikato, it is considered that this approach has led to a more adequate and realistic view of the instructional system by the teaching staff and to better physics instruction.