FORMATIVE EVALUATION IN UNIVERSITY SCIENCE COURSES

by

D. J. BOUD

(In partial fulfilment of the requirements of the PhD degree, University of Surrey, October 1974)
SUMMARY

The formative evaluation of university science courses aims to provide information about such courses that can be used to facilitate their improvement. An account is given of examples of the formative evaluation of three first and second year undergraduate courses in physics and biology. These case studies provide the basis for discussion of some key issues in formative evaluation. They are the role of the evaluator, the relationship between evaluators and university teachers, and the kind of evidence that is valid for course improvement.

The case studies are also examined with respect to the style of evaluation that was the most productive. Although the individual aims of each study were met, deficiencies in the evaluative approach were discernible in some cases. An examination of these shortcomings led to a proposal for a style of evaluation called 'supportive evaluation'. This is described in terms of the relationship between the evaluator and teacher and the responsibilities of each party.
"In the course of the year he asked the men each to write some word of suggestion, if he were so inclined, for improvement in the method with which the course was conducted; and, if I remember rightly, there were not a few respectful suggestions that too much time was allowed to the few wrangling disputants. In a pretty full and varied experience of lecture-rooms at home and abroad I cannot recall another where the class was asked to criticize the methods of the lecturer!"

Dickinson S. Miller of William James, 1842-1910
<table>
<thead>
<tr>
<th>Part I</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>1.1</td>
<td>Aims</td>
</tr>
<tr>
<td>1.2</td>
<td>The context</td>
</tr>
<tr>
<td>1.21</td>
<td>The immediate context</td>
</tr>
<tr>
<td>1.22</td>
<td>The broader context</td>
</tr>
<tr>
<td>1.3</td>
<td>Rationale</td>
</tr>
<tr>
<td>1.4</td>
<td>Overview</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2.</th>
<th>The Literature of Course Evaluation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Some approaches to the conceptualisation of courses</td>
<td>10</td>
</tr>
<tr>
<td>2.21</td>
<td>The need to conceptualise</td>
<td>10</td>
</tr>
<tr>
<td>2.22</td>
<td>Three schemes</td>
<td>11</td>
</tr>
<tr>
<td>2.221</td>
<td>The systems approach</td>
<td>11</td>
</tr>
<tr>
<td>2.222</td>
<td>Curriculum development approaches</td>
<td>16</td>
</tr>
<tr>
<td>2.223</td>
<td>Curriculum evaluation models</td>
<td>19</td>
</tr>
<tr>
<td>2.2231</td>
<td>Stake's model</td>
<td>19</td>
</tr>
<tr>
<td>2.2232</td>
<td>Taylor and Maguire's model</td>
<td>22</td>
</tr>
<tr>
<td>2.224</td>
<td>A review of the three approaches</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Two aspects of the approaches</td>
<td>28</td>
</tr>
<tr>
<td>2.31</td>
<td>Aims and objectives</td>
<td>28</td>
</tr>
<tr>
<td>2.311</td>
<td>Values, aims and objectives</td>
<td>29</td>
</tr>
<tr>
<td>2.312</td>
<td>Formulation of objectives: problems and issues</td>
<td>30</td>
</tr>
<tr>
<td>2.32</td>
<td>Evaluation</td>
<td>40</td>
</tr>
<tr>
<td>2.321</td>
<td>What is evaluation?</td>
<td>40</td>
</tr>
<tr>
<td>2.322</td>
<td>Judgment and evaluation</td>
<td>42</td>
</tr>
<tr>
<td>2.323</td>
<td>Types of evaluation activity</td>
<td>43</td>
</tr>
<tr>
<td>2.324</td>
<td>Evaluation and hypothesis testing</td>
<td>44</td>
</tr>
<tr>
<td>2.325</td>
<td>Non-prescriptive evaluations</td>
<td>46</td>
</tr>
<tr>
<td>4.102</td>
<td>Validity</td>
<td>92</td>
</tr>
<tr>
<td>4.103</td>
<td>Other evidence of reliability</td>
<td>93</td>
</tr>
<tr>
<td>4.10(3)1</td>
<td>Course A</td>
<td>93</td>
</tr>
<tr>
<td>4.10(3)2</td>
<td>Course B</td>
<td>94</td>
</tr>
<tr>
<td>4.11</td>
<td>Conclusion</td>
<td>95</td>
</tr>
</tbody>
</table>

| 5. | An Introductory Biology Lecture Course | |
| 5.1 | Introduction | 97 |
| 5.2 | The course | 97 |
| 5.3 | Specific constraints | 99 |
| 5.4 | Rationale for choice | 100 |
| 5.5 | Pilot background study | 101 |
| 5.51 | Student aims | 101 |
| 5.6 | Approach adopted | 102 |
| 5.61 | Aims questionnaire | 103 |
| 5.62 | Pre- and post-attitude measure | 104 |
| 5.63 | Ancillary measures | 105 |
| 5.7 | Results and analysis | 105 |
| 5.71 | The aims questionnaire | 105 |
| 5.72 | Pre-and post-attitude measures | 110 |
| 5.73 | Ancillary measures | 112 |
| 5.8 | Discussion of results | 114 |
| 5.9 | Outcomes | 116 |
| 5.10 | Statistical footnote | 118 |
| 5.101 | Reliability | 118 |
| 5.102 | Validity | 119 |
| 5.11 | Conclusion | 119 |

| 6. | Quantum Mechanics and a New Teaching Method | |
| 6.1 | Introduction | 121 |
| 6.2 | The course | 121 |
| 6.3 | Background to the course | 124 |
| 6.31 | The earlier course | 124 |
| 6.32 | The influence of Keller Plan courses | 126 |
| 6.4 | Specific constraints | 128 |
III  Quantum Mechanics: Course Feedback Sheet, 1971
IV  Quantum Mechanics: Questionnaire, 1971
V   Hypotheses about the Keller Plan Method of Learning, 1973
PART I

Chapter One : Introduction

Chapter Two : The Literature of Course Evaluation

Chapter Three : Some Issues in Evaluation Relationships
1. INTRODUCTION

1.1 Aims

The purpose of the work to be described was to investigate methods of formative course evaluation for university science courses. This broad aim was interpreted in the light of the availability of courses to work with and more specific aims for this study were adopted. They are:

- to examine three examples of courses and to develop methods of evaluation appropriate for their improvement.
- to treat these evaluations as case studies in formative course evaluation
- to examine them in the light of some issues in evaluation which have been voiced by other practitioners in this field
- to discuss the problems which emerge from the case studies, especially those concerned with the role of the evaluator and the strategies which were facilitative of course improvement
- to review the case studies and the evaluative role adopted with respect to criteria which have been proposed for effective evaluations.

1.2 The Context

1.21 The Immediate Context

All of the work described, except where specifically stated, was carried out at the University of Surrey. It was based in the Institute for Educational Technology as part of the Institute's programme of research and development in
university teaching which had the aim 'to improve teaching and learning within the University'. The setting was one which was predominantly development rather than research orientated. Emphasis was placed on work that had tangible, relatively short-term outcomes and which resulted in a perceived change in particular areas of university teaching.

The research was thus highly problem centered and was done under the same or similar constraints as real-time evaluation of university courses. No attempt was made to alter the courses that were being examined to fit them to the evaluation methodology. The courses were taken, with all their constraints and inconsistencies, as the prime object of the research.

However, early decisions were made about the orientation of the evaluation. Many perspectives could have been chosen. For example, an attempt at an external objective evaluation could have been made, or a viewpoint of an educational administrator could have been adopted. The viewpoint chosen was that of the university teacher who was responsible for the course. He was seen as the chief client of the evaluator. This was accepted as a real-life constraint - it was likely to be the viewpoint of people working in the Institute towards courses they were working with.

The choice of viewpoint had major implications for the approach that was adopted. It meant that the first criteria to be satisfied were those of the teacher and the freedoms and limitations which he allowed. Traditional research
criteria were secondary. In committed research of this kind this is not an inevitable limitation on an otherwise free research, but a part of the phenomenon that is being investigated. Other perspectives could be chosen, but they would not be able to fulfil the aims of this study.

1.22 The Broader Context

Although the work was specifically problem centered, it addressed problems that were within the broader range of evaluative research. Using the terminology of Scriven (114) the evaluation can be described as formative, that is, it was developmental in nature and was designed to help improve the courses that were being examined, rather than summative, it was not concerned with helping to pass judgements on a final product, the courses were seen as evolving. It also concentrated on performance (50) or pay-off evaluation (114) rather than intrinsic evaluation. That is, it is based on measures of student performance, what students were or were not able to do, or how they did or did not see things, as distinct from measures of the intrinsic content of the course: the content and organization of subject matter within a lecture, for example.

The restriction to the area of formative, performance evaluation requires some justification. In particular, the exclusion of intrinsic evaluation does not follow from the simple definition of the aims of the research. However, this restriction is plausible if the constraints of the research are examined. The participants in the evaluation were a teacher, or teachers; the students enrolled on a particular course; and the single evaluator. In this situation the
teacher(s) acted as the sole subject experts and the evaluator acted as the determiner of educational objectives and gatherer of performance data. A notion of intrinsic evaluation could not be introduced into this situation without either introducing more participants, which was outside the resources of the project at that time, or the evaluator acting as a subject matter specialist. This second alternative was rejected for two reasons. Firstly, the evaluator was seen as a research student in another subject and therefore did not possess, or want to possess, any credibility as a subject expert. Secondly, such a role would have disrupted the role of the evaluator with respect to the other tasks he had to perform. This aspect is discussed at greater length in Section 3.4.

1.3 Rationale

The project was originally conceived as one which would develop specific course evaluation methods which could be applied in a variety of courses with relatively little modification. This conception was soon rejected as being unrealistic, or at least premature. There were many questions to be asked and answered about the role of evaluation for improvement in courses and the reactions of teachers to this, before intensive development of any particular techniques were initiated. The project was therefore restated as one which would investigate how evaluation for improvement can be developed in a particular course in conjunction with the teacher in such a way that it would be fully accepted by the teacher and would lead to observable improvements in the course being studied. This
problem is a realistic one. Many universities are setting up units to improve teaching and learning, called Centres or Institutes of Educational Technology, University Teaching Methods Units, etc., and such units are searching for ways in which they can work with university faculty. Much of the work is in course development, particularly in new courses and it is especially important that in these areas evaluation is introduced at the formative stage. In some institutions, like the Open University, procedures are laid down and co-operation with its Institute of Educational Technology is prescribed for the faculty. In most institutions this situation does not exist: most collaborations between faculty and educational technologists are voluntary, and in these cases it is essential that educational technologists develop ways of working which are acceptable to the faculty and which are seen to be productive.

It is in the light of these considerations that the present work developed. In the early stages university teachers were approached to take part and the exercise was seen as more of a pure research activity (see Chapter 4), but later the activity was recognised to be more of a formative evaluation, one in which the teacher and the evaluator worked together (see Chapter 6).

1.4 Overview

The report represents a retrospective account of some evaluation activities which, when conducted, were construed by the author in ways different to those that have now been adopted. The empirical studies reported
in Part II (Chapters 4, 5 and 6) were completed in the years 1969-1972. These form the basis for a discussion of problems in formative evaluation in higher education which starts in Part I and concludes in Part III.

The format which is adopted here is one that takes the three case-studies in evaluation as the core of the thesis. They are examples of evaluations which in retrospect, although incomplete, did fulfil the objectives which were set for them at the time of execution. They are presented in their original form so that they can be examined from the perspective that has arisen from them. They form the raw material for an examination of relationships between the evaluator and his client and between the evaluator and the environment in which he finds himself. These are discussed in Chapter 3 "Some Issues in Evaluation Relationships" and further in Chapter 7 "Judgements and Conclusions".

In Part I, Chapter 2 is a survey of the literature concerning formative course evaluation, particularly in the context of course development. It concentrates on the perspective that formed the basis for the case-studies. Chapter 3 outlines some of the issues that arose during the total period of the study. In it stress is given to those issues related to the role of the evaluator, the constraints which act upon him in the university course context, and the relationships between evaluator and teacher. It concludes with a discussion of the kind of data that should constitute evidence in evaluation and with a section presenting the perceived bias of the author.
Part II consists of the empirical studies:
Chapter 4 "An Introductory Physics Laboratory Course",
Chapter 5 "An Introductory Biology Lecture Course" and
Chapter 6 "Quantum Mechanics and a new teaching method".
These are self-contained reports of three investigations
of first and second year university courses. They are
presented in similar formats to allow comparisons to
be made. They are described in terms of the issues
that the evaluator was conscious of at the time of the
study. This is to enable the development of the present
position on evaluation to be understood.

Part III is the concluding section. Chapter 7 reviews
the case studies and makes judgements about their worth
in terms of criteria for good evaluations that have been
developed by Stufflebeam and his colleagues. It continues
by examining the style of these evaluations and in the light
of the differing successes of these studies proposes an
evaluation model which the author considers appropriate for
the situations in which the research presented was conducted.
Finally, conclusions are drawn about the direction that
formative evaluation should take in university science courses
and where effort should be concentrated in this area.
2. THE LITERATURE OF COURSE EVALUATION

2.1 Introduction

The aim of this chapter is to provide a perspective for viewing the evaluation studies reported in Chapters 4, 5 and 6 and to describe a variety of conceptual frameworks suggested by workers mainly in the United States which are useful for understanding processes of course evaluation. The emphasis will be on approaches for conceptualising courses to assist with evaluation, rather than, for example, conceptualisation for the purposes of course design or for guiding the development of instructional methods.

At the present point in time it is only possible to present descriptive models for course evaluation (122), not theories. The aim of these models as presented here is to give a systematic view of the field of study and to suggest guidelines for evaluation procedures (62). Three models will be described. Firstly, that derived from the systems approach to education based on the application of systems analysis ideas to educational problems (126); secondly, one derived from the field of curriculum development, based largely on the work of Tyler and his colleagues (138) and finally, particular curriculum evaluation models proposed primarily for the purposes of guiding evaluation (122,134).

The three approaches will be examined with respect to their utility in guiding course evaluation procedures and to their ability to cope with the many constraints that are inevitably present in any given situation. Two
aspects of these approaches will be focused on. Firstly, the role that aims and objectives play within each model and secondly, the particular role that evaluation takes in each.

The final section of the chapter will deal with the extent to which course evaluation procedures have been developed in higher education in this country and also with the differences to be found in the approach adopted in the studies reported later and those found in most present evaluation exercises.

A large gap has been recognised by many authors in this field (125) between the theoretical schemes proposed and the practice of course evaluators. Often evaluators have tended to rationalise their practice by subsequent resort to models or have proposed models without empirical tests of their applicability. However, there are signs that some models are being systematically explored and probed for their utility in practical course evaluation (102). It is difficult to recognise the true worth of these approaches until they have been tried in circumstances that are at least in some way similar to those found in British Universities. It is a large step, for example, to transfer ideas that have been tried in elementary schools in California (98,103)* to a Technological university in England. Nevertheless, attention in this chapter will be focussed on the major attributes and limitations of some of these models.

* much of the work on the application of systems analysis to education has taken place in American elementary schools by organisations such as the Systems Development Corporation of Santa Monica, California.
2.2 Some Approaches to the Conceptualisation of Courses

2.21 The Need to Conceptualise

The processes that take place in even the simplest course in a school or university are highly complex. A huge number of interactions take place between staff and students. The expectations of the member of staff in terms of what students should learn in a single lecture alone would normally cover several pages. Indeed if they were stated as some specialists have proposed, in behavioural terms, they would be considerably longer. How much more difficult is it then for a complete course to be described, when one would have to take into account student learning, attitudes, organisation, methods, assessment, etc.

There is a need for schemes (137) whereby the activities of a course can be described in a systematic manner, as something more than a list of course content. The basic criteria for such a scheme would be that it should consider all of those elements that are generally recognised as forming a part of what is known as the educational or instructional processes of a course. For the purposes of improvement it should focus on those aspects that were amenable to change and variation within the normal limits of time and resources. That is, an important practical criterion for such a description would be that it took particular cognisance of those aspects of the course that could be altered in such a way that would lead to some recognisable improvement in the final outcomes.
It is this 'description as a basis for improvement' that forms the rationale for the choice of the models described in the following sections. They are conceptualisations of courses or curricula that will be discussed in terms of their utility for evaluation for improvement (40). It is analogous to the use of models in the physical sciences (87) where a conceptual scheme can provide a theoretical background to the empirical investigations which can suggest possible areas of study, so giving direction to those aspects of a given problem, be it a physical phenomenon or a university course, which should be studied.

Also, and just as important, a scheme can provide a means for communication. It can suggest a terminology that is based on something more than ad hoc probing of the entity being studied. It can enable evaluators to communicate with the teachers with whom they work, and with each other.

2.22 Three Schemes

In the three examples of conceptual schemes which follow, discussion will centre around those elements in the models that have a direct relationship with evaluative processes and the facilitation of such processes, and that can most adequately describe the full richness of educational activities which they strive to represent.

2.221 The Systems Approach

The application of systems approach methods to education arose subsequent to their use (primarily by the military and
later by engineers and by management) for the analysis of complex organisational problems (13). Mathematical techniques derived from operations analysis and systems analysis were used over an increasingly wide area for problems involving human beings as well as machinery (89). This development was brought about by the need to investigate complicated systems involving many components and inter-relationships in a systematic manner and to change these systems in such a way that the outcomes were made optimally efficient. Meals (89) characterises systems analysis as "The application of scientific methods and tools to the prediction and comparison of the values, effectiveness, and costs of a set of alternative courses of action involving man-machine systems".

It has only been in the last five years, however, that these techniques have seriously been applied to educational problems (38), due largely to the sponsorship of U.S. Department of Defense (13).

Various system models have been proposed for a wide variety of specific educational processes from timetabling to counselling (103). The systems model discussed here has been proposed by Lehmann for the development of an educational programme (77). It will be considered in terms of the outlook of Brattan (25) towards the systems approach to education.

Lehmann (77) reports the work of project ARISTOTLE in developing a systems model with eight steps. He adopts the premise that "the systems approach does provide an orderly process for developing a solution, a process which is structured to minimise prejudicial preconceived notions..."
and maximise the objectivity required to arrive at a scientifically correct answer ". He also states that the approach that he presents "is what we have called in the past 'the scientific method'". He does not, however, offer any evidence to support this assertion.

The eight components are as follows:

NEED: a statement of the real problem being faced by the society under consideration - that statement of a problem which initiates consideration of an education/training system as a potential solution.

OBJECTIVES: the determination and specification of the terminal capability desired of students after having successfully completed a learning experience.

CONSTRAINTS: those real-world limiting conditions which must be satisfied by any acceptable system designed to attain the educational objectives.

ALTERNATIVES: the generation of candidate systems which could achieve the objectives.

SELECTION: the systematic evaluation of all alternatives in terms of objectives and constraints to select one which is considered the most desirable.

IMPLEMENTATION: the first adoption of the selected alternative to meet the specified objective.
EVALUATION: the determination of the conformance or discrepancy between all of the objectives initially specified and the performance that was actually obtained.

MODIFICATION: the process of modifying the designed learning system based on deficiencies in meeting the objectives as determined through evaluation.

Lehmann emphasises the crucial step as that of the specification of objectives, for without this the subsequent steps become valueless. Indeed, what is necessary in this model is that the objectives be specified in behavioural or operational terms, i.e. in such a way that it can be unambiguously determined whether or not they have been achieved. When this has been done evaluation is seen as the process whereby the objectives are checked to see if each one has been achieved or not. Thus, in the systems approach described evaluation rests upon the close matching of evaluation procedures with the clearly defined objectives which have been established prior to the course.

The systems approach does not however remove all subjective elements from the design and implementation processes. Lehmann himself points out that the SELECTION component must be performed scientifically and should not involve bias. However, he goes on to suggest that "horse sense, experience, and good calm judgement should be introduced". And it is not only in SELECTION that this procedure is required. It seems just as necessary in EVALUATION where the same element of 'judgement' is required to match evaluation to objectives.
So in the particular model described here we have a systematic approach to design and decision-making concerning courses, but it is one which inevitably has both explicitly and implicitly incorporated elements of subjective judgement - an act which can produce a source of bias. What the approach does make clear is the rationale of systems approaches, that is, it demonstrates a process of conceptualising courses which emphasises the need to make explicit more of the decision-making processes in course design and evaluation. It also presents a scheme to facilitate the explication of these processes so that they may be carried out in a more systematic manner. Indeed, Bratten (25) adopts this approach when he defines a systems approach to education as "an attitude or conviction that educational processes...... should be viewed as systems ". He believes that "as systems, educational processes have ascertainable and measurable products; that the essential internal components can be described; and that the resources necessary for the functional means to attain the measurable ends can be specifically known ". It is this belief that all educational processes can be objectively described and objectively measured which characterises the systems approach to the solution of educational problems. It is this contention that objectivity and rationality are foremost that perhaps accounts for the relative lack of examples of the application of these opinions in real-life educational contexts.

This difficulty is aptly described by Ruth Beard in her monograph on "Research in Teaching Methods in Higher Education " (6) when she quotes, "it was striking to find that university graduates, with sometimes years of research experience, were unable to apply to the problems of teaching
the same methods of scientific inquiry they would, presumably, use in their own discipline ".

Although applying systems approach techniques in toto to real problems may present great difficulties, this does not mean to say that such a stance is not useful. Neil (93) in describing the development of courses at the Open University points out that although the method that they have adopted could not strictly speaking be described as a systems design, it does, however, contain sufficient elements and attitudes from this kind of approach to warrant it being called systematic. It is this 'systematic' nature that provides a language for the rational discussion of course design and evaluation.

2.222 Curriculum Development Approaches

One example of an approach to curriculum evaluation will be given here: that of the classic Tyler curriculum development methodology. Historically curriculum development schemes pre-date by almost 30 years the systems approach methodology outlined in the previous section: the pioneering work in this area being centred around Ralph Tyler in the early 1930's onwards.

The approach is encapsulated in the rationale proposed by Tyler in 1949. He begins by identifying four fundamental questions which must be answered in developing any curriculum. They are:

"1. What educational purposes should the school seek to attain?

2. What educational experiences can be provided that are likely to attain these purposes?"
3. How can these educational experiences be effectively organised?

4. How can we determine whether these purposes are being attained? " (138)

He goes on to suggest methods for studying these questions, thereby constituting a rationale by which problems of curriculum and instruction may be examined.

Attention is drawn to the totality of the educational experiences. The four questions can be thought of as being a less specific formulation of the systems methodology. The difference in practice between the two lies in the way that they diverge from this point - the systems approach takes an increasingly microscopic and analytical view, whereas the Tylerian model takes a broad perspective until quite late in the process. Much greater emphasis is placed in this approach on the scanning for appropriate educational objectives (purposes), than on analytical procedures for the specification, attainment and rigorous testing of them. Objectives, for example, encompassing studies of contemporary life outside school, philosophy and psychology of learning can be selected and these emphasise Tyler's strategy of not specifying too closely too soon.

Tyler was one of the first to stress the close tie between educational objectives as a basis for instruction and as a basis for evaluation, and he proposed some guidelines for the formulation of these objectives (90,119). The elements of a currently accepted (18), objectives-based, Tylerian model as related to evaluation are as follows:
"1. Formulate objectives. Determine the broad goals of the program.

2. Classify objectives. Develop a typology of objectives so an economy of thought and action can be achieved.

3. Define objectives in behavioral terms. This guideline is basic to the model. 'Modern' approaches to evaluation rest heavily on the specific, behavioral statement of objectives.

4. Suggest situations in which achievement of objectives will be demonstrated and a minimum level of achievement requested.

5. Select or develop appraisal techniques (standardised tests, ad hoc tests, questionnaires, etc.)

6. Measure student performance and compare performance data with behaviorally stated objectives." (18)

The chief characteristics of this evaluation model is that it places high priority on student behaviour. Objectives are stated in behavioural terms (see 2.211), and information is gathered to test whether or not the desired behaviours have been achieved. It is a model centred around the ends of instruction, rather than the means. It has been criticised for this emphasis. Critics (46, 84) claim that it is not possible to measure the ends easily at all. Often the ends of instruction are embodied in a personal life-style which is almost impossible to measure and to relate to the original educational experience. Also this information is long term and so is of little help to the educator for short-term course improvement. By concentrating on a broader perspective than systems models it meets the criticism of narrowness of intent, but fails on the likely measurability of outcomes.

It must be noted in support of this model, however, that it retains a certain utility of implementation which has caused
it to be the basis for some of the currently more systematic curriculum development exercises in this country (133).

In summary, the Tylerian evaluation model is dependent on the prescription of student behaviour. Its strengths and weaknesses lie in its ability to achieve this prescription and test it effectively.

2.223 Curriculum Evaluation Models

Although the curriculum development models include a component of evaluation, a few models have been proposed which concentrate almost entirely on the evaluation sector of the course or curriculum. They are those of Stake (122), basing his ideas on Scriven (114), and Taylor and Maguire (134).

2.2231 Stake's Model

This is based upon a conception of evaluation which stresses evaluation as a process of determining the worth of an educational activity (for further discussion of this see 2.32). It has two main components - a description matrix and a judgement matrix, each of which are subdivided into cells which represent the basic components of the evaluation task (see Fig. 2.1) The data to be collected by the evaluator is of three kinds: antecedent, representing events and activities prior to the teaching and learning under study; transaction, representing encounters between the people (students, teachers, counsellors, etc.) and the materials and activities of the
Figure 1. A layout of statements and data to be collected by the evaluator of an educational program.

Figure 2. A representation of the processing of descriptive data.

Fig. 2.1 Stake's Model

Figure 3. A representation of the process of judging the merit of an educational program.
learning situation; and outcome, representing the products of any kind that appear.

There are two further components of the description matrix - intents and observations. Stake regards 'goals', 'objectives' and 'intents' to be synonymous, but he uses the term intents to distinguish his category from those used in other models, cf. systems approach which equate 'goals' and 'objectives' with intended student outcomes. Intents include everything which is planned for, effects which are desired, those which are hoped for, those which are anticipated, and those that are feared. Also into this category come the intents of students. No criteria for acceptability are prescribed "taxonomic, mechanistic, humanistic, even scriptural - any mixture of goal statements is acceptable as part of the evaluation picture" (122).

Observations are descriptions of surroundings and events and the subsequent consequences. These will include intended student outcomes, but will not be restricted to them. They cover all aspects of the activity which answer the question - "what happened?".

The judgement matrix has two similar components - general standards and specific judgements. Stake holds that the evaluator is responsible for making known what standards of excellence are held by various reference groups. In addition to including the standards, the evaluator should record the weights various judges assign to the set of standards. The judgement matrix represents an integration of descriptive data and judgement in a form which facilitates decision-making. (101)
The final component of Stake's model is the rationale. This is the theoretical basis of the course and it should provide one basis for evaluating intents, i.e. whether the plan proposed for the course is a logical step in the implementation of the general purposes of the course. It also gives a guide to selecting groups or persons to pass judgement on various aspects of the course.

Stake finally places emphasis on a methodology of evaluation which consists of the evaluator finding the logical contingencies or relationships between antecedents, transactions and outcomes; and finding the congruence or matching between Intents and Observations. That is, do the transactions follow from the antecedents, and the outcomes from the transactions? and does what the course designers intend correspond to what actually happens?

The utility of this model is difficult to assess. It has only been applied in two known cases, that of Steele (127) and PEEP (102), which is the major example. The Steele study is an unpublished doctoral dissertation, and the PEEP study has yet to be written up. The model is at present a theoretical one, and so initially it can only be assessed as such.

It is not prescriptive in a detailed sense. It does not give detailed guidelines to the path that the evaluator should follow. It does indicate the main features of a study that should be looked into, but it gives little indication of priorities to be followed. This can be regarded as a positive benefit compared to some highly analytic models.
which pre-specify too closely (see 2.221); although in practice this is a disadvantage. Methods have yet to be developed for handling the analysis of contingencies and congruities of educational programmes.

Another difficulty lies in the judgement side of the model. It is a current debate in the literature of evaluation whether the evaluator should or should not be involved in judgement problems (40, 114). Stake's model is constructed with judgement in mind and a considerable part of his effort has gone into integrating this aspect of evaluation with the traditional side of descriptive and normative evaluation. If, as we shall discuss in 2.32, one adopts a view of evaluation as an aid to decision-making, it is necessary in some part to arrange evaluative data in a form which allows judgements to be made even if it is not the evaluator's responsibility to make these judgements. This aspect Stake allows for. He accepts that evaluator and educator may not be two distinct roles and that what he is concerned with is that part of educational activity that can be labelled 'evaluation'. The role of the person identified as 'evaluator' remains to be discussed (see section 3.4).

**2.2232**

Taylor and Maguire's Model

In many ways this can be thought of as descending from Tylerian curriculum development models whilst at the same time recognising the non-sequential nature of evaluation activities and the importance of judgement, as in Stake. Taylor and Maguire (134) suggest that the uses of their model lie in suggesting variables and relationships to be investigated in the course of evaluation, and that it has pedagogical merit both in terms of schematizing the evaluative process and in
indicating tasks for which evaluators should be trained.

The elements of the model are as follows: (see Fig. 2.2)

**Institutional Press**: "those societal and professional pressures that lead to the statement of broad-category objectives which define the relationship between school and society", i.e. from parents, academics, politicians, pressure groups, etc.

**Broad Objectives**, B: these are derived from the institutional press. They are insufficiently precise to be of great value in curriculum development, except as a broad guide.

**Interpretations**, I: it is the responsibility of curriculum planners to interpret the broad objectives into specific descriptions of behaviours that they would anticipate a student to exhibit consequent to exposure to that facet of the educational program to which the objective is orientated.

**Strategies**, S: curriculum developers translate the interpreted objectives into possible classroom strategies. Each interpretation gives rise to two components of strategy - Elicitations, E, and Presentations, P. These are respectively the activities of the program, and the substantive content of, for example, information imparted.

**Outcomes**, O: this covers two types of behaviour that the student might exhibit, namely, outcomes on performance specifically oriented toward the school setting, and
Societal Press

Judgmental Stages

Broad Objectives $B_1, B_2, ..., B_k$

Interpretations $I_{1j}, I_{2j}, ..., I_{kj}$

Strategies $E_{1,jh}, E_{2,jh}, ..., E_{k,jh}$

Outcomes $P_{1,jh}, P_{2,jh}, ..., P_{k,jh}$

Fig. 2.2  Taylor and Maguire's Model
outcomes that are generalised and not apparently related directly to the curriculum.

The preceding description is essentially that of a curriculum development process, and has great similarity to that of Tyler (2.222). However, Tayler and Maguire include specific statements of where and how the evaluator should direct his attention.

The task of the evaluator up to Stage I is to obtain information which would enable him to recognise additions and deletions to the list of broad objectives that have occurred between the statement of a broad objective and its translation into behavioural terms. These may include statements from representatives of the institutional pressure groups on the extent to which the behaviour statements reflect the essence of their original intention. At Stages S and O the evaluator is responsible for obtaining feedback information from the classroom situation and from student performance for those concerned with the development of the curriculum so that adjustments can be made to the strategies, where necessary, or, even additions or deletions, from the list of behaviours I.

In addition to these activities, the evaluator must maintain a judgement role. At each stage the evaluator should elicit the services of external 'experts' to judge the worth and feasibility of the activities of the curriculum developers. The stages of judgement are:

A relative worth of broad objectives, from 'social philosophers' and/or empirical surveys
B the utility and appropriateness of interpretations at two levels - social, i.e. whether it is the proper concern of the school, and administrative, i.e. does it have high enough priority to be included in the programme.

C the efficiency and adequacy of presentation and elicitation, potential effects on student outcomes, and the correspondence of the strategies to the interpretations.

D the assessment of the goodness of fit of the observed outcomes to the criterial behaviours listed as objectives in I.

The difficulties of acceptibility of this model are very similar to those related to Stake's model. Firstly, the lack of studies based on this model presents the problem of whether or not the theoretical model can be implemented. This remains to be seen. Secondly, the issue of the role of the evaluator in judgement of an educational activity emerges again. Taylor and Maguire suggest possible sources for judgement data, but they do not make explicit the specific relationships between the 'educators', 'evaluators' and 'experts', which is a potential source of great difficulty in practice. However, Taylor and Maguire do present a model which is more specific than that of Stake. In doing so they have committed themselves to a particular philosophy of education, namely that embodied in the essential use of behavioural objectives (2.312) and this means that the model must stand or fall together with behavioural objectives.
All three approaches have in common the desire to guide the evaluation process. However, the systems approach and to a lesser extent the Tylerian approach see evaluation as a totally integrated part of the curriculum development process and thereby limit the range of possible evaluation activities that can be adopted to those related directly to the planned curriculum. The systemists see evaluation solely in terms of intended student outcomes, and Tyler sees evaluation in terms of intended and unintended student outcomes, whereas the curriculum evaluation approaches adopt a wider perspective which focuses evaluation on all aspects of the curriculum process.

No one model encompasses the variety of possible curriculum evaluation processes. This is made explicit by all but the systemists. The Stake approach could be considered to be fairly all-encompassing except for its dearth of specific procedures. However, most systematic evaluation techniques could be fitted in without much distortion.

The two curriculum evaluation approaches largely adopt a teacher/curriculum developer view of education. The possibility exists with Stake of adopting an alternative framework so that, for example, a course might be evaluated from a completely different perspective to that of its designers. This is theoretically possible with Taylor and Maguire, but the model is loaded so that the crucial steps of evaluation are left to the curriculum developers.
With Scriven and Stake's emphasis on the evaluator as the specifier of objectives, some degree of independence from the original course designers is achieved.

Firstly, what problems are not accounted for in these approaches? The most basic difficulty of evaluation is that it usually takes place within a rapidly changing context, it is dynamic. It is rare for a curriculum to retain the same aims and objectives, or even standards throughout its active life. As experience of the programme in operation is gained, so are the objectives modified. This is a problem which has received little attention from proponents of models of evaluation. It is normally recognised as an issue, but it is often not built in as a factor. For example, Taylor and Maguire present a model as a flow process whilst at the same time giving a caution that their model should not be seen as sequential. It may be that models are often a rationalisation of experience after the event, which is not a bad thing in itself. The difficulties arise when such examples are used as a guide to action.

However, a more fundamental criticism remains. None of these approaches accept the mutuality of the course or curriculum situation. That is, that all education is concerned with exchanges between two groups of people, labelled as either staff, curriculum developers, text-book writers, or, as pupils or students. These exchanges will be perceived differently by the different groups and by different individuals within a group. Whilst adopting the perspective of members of one of these groups, or, more specifically, one sub-group called evaluators, the evaluation models can only represent the object of evaluation -
the educational experience - from one point of view. It is not enough to collect data and receive criticism from the other parties if the conceptual framework in which the interpretation is set is rooted in a totally different context. Students' viewpoints, for example, are recognised in so far as their perceptions of the situation approach and are reinforced or negated directly by their teachers. If their conception of the experience is substantially different then this cannot be taken into account.

It would seem that this would be a most fruitful area for future development of evaluation models. It is not intended to propose a detailed, alternative framework to take into account these criticisms, however, they will be recognised in the interpretation of the work presented in later chapters.

2.3 Two Aspects of the Approaches

Two common elements are found in all of the evaluation processes: those called aims, objectives, goals or outcomes and those that fall within the word-evaluation. The following sections will review more closely the differing views of aims and objectives, and evaluation that have been proposed and will concentrate on the definitions that have been adopted in the rest of this report.

2.3.1 Aims and Objectives

It has commonly been accepted (114) that it is not possible to prescribe evaluation strategies without a
reasonably detailed knowledge of the goals of a given programme. The form of specification of aims and objectives, however, has been the subject of much debate (106,124). Three aspects of this are:

(i) for what purpose are aims and objectives specified?

(ii) at what level of generality are they formulated?

(iii) what role do they perform in evaluation?

The following sections will discuss these questions and relate them to the different evaluation models mentioned earlier. It will end with a section describing the standpoint on aims and objectives taken in this report.

2.311 Values, Aims and Objectives

In practice, the difference between aims and objectives is often slight. On the whole aims are used to indicate more abstract, general and value-oriented goals, whereas objectives are used to indicate more specific and descriptive goals (107). Goodlad (55) has devised a conceptual system for the curriculum which clarifies these distinctions and relates them to values. He maintains that it is necessary to take one's ultimate source of educational aims and objectives from values that are given high regard by society and which are supported by the fund of knowledge held by that society and by its conventional wisdom. From these can be derived educational aims, which in turn give rise to educational objectives and learning opportunities, so
that in rational curriculum planning one would move from the general to the specific formulations and one could finally define objectives with great specificity in organising learning programmes for specific individuals or groups. Although Goodlad continues to make specific recommendations about the procedures and methods that can be adopted at each stage, it is sufficient to give this as an example of the hierarchy of values, aims and objectives that is currently accepted. Similar schemes, less refined have been put forward by Stake (123), Pring (107) and other educational philosophers.

This hierarchy draws attention to the process of selecting aims and objectives. It is this area that has probably been the subject of the greatest discussion in the evaluation literature. It centres around the specific problems of how objectives can be formulated, and for what purposes they can be used.

2.312 Formulation of Objectives: Problems and Issues

Implicit in some curriculum models is the need to formulate objectives in operational terms. This has been accepted for both the systematic approach (111) and for the Tylerian approach (18). It is necessary in these cases to write objectives in such a way that test items can be derived from them which will directly measure the intended outcomes on the part of the student. Objectives specified in this manner are commonly known as behavioural objectives.

Eisner (44) has formulated three characteristics necessary for preparing a useful statement of objectives.
These are:

1. Educational objectives should describe student behaviour, not teacher behaviour - they should describe how students are to perform subsequent to educational experiences.

2. They should describe both the behaviour to be displayed and the context in which the behaviour is to occur, i.e. critical thinking in maths.

3. They should be stated at a level of specificity that makes it possible to recognise the behaviour should it be displayed, e.g. not critical thinking.

Objectives meeting such criteria, he claims, facilitate a number of subsequent functions:

1. They give direction to curriculum planning.

2. They provide criteria for selecting content and organising curriculum activities.

3. They provide cues for formulating evaluation procedures in as much as evaluation should proceed from specifications set forth by objectives.

Maguire (86) reviewed some purposes served by stating educational objectives:

1. to direct ongoing classroom instruction
2. to guide in the selection of content
3. to aid in the evaluation of student progress
4. to direct course evaluation activities
5. to aid in the development of new courses
However, he points out that "the purposes served by objectives may dictate the form of their statement" and that "it is important that the rationale for decisions at various levels be made explicit, so that conflicts between initiators and recipients can be resolved, thus maximising the chance of achieving the broad objectives that are deemed significant by the society".

Krathwohl (70) points out that it is necessary to analyse objectives to several levels of specificity depending upon how it is intended to use them. He proposes three levels:

First level: broad and general statements - general goals of courses, etc.

Second level: behavioural objectives useful for specifying the goals of an instructional unit, a course, or a sequence of courses.

Third level: needed to create instructional materials - the operational embodiment of one particular route to the achievement of a curriculum planned at the second and more abstract level.

He argues that the first two levels are required:

1. as a guide to the educational process
2. because not all objectives lend themselves to specification at the third level
3. so that we can continually examine their interrelation to one another. It is easier to obtain concensus at more abstract levels
4. as there are many possible routes from the first to the third level. Different routes can be regarded as sub-objectives which need evaluation to find the best one. The general objectives provide reference points for this activity.

Scriven (114) has suggested a similar classification scheme which has been advocated in this country by MacKenzie et al (85) and discussed by Stones (129). He identifies three levels of description for educational objectives:

1. the conceptual level, relatively abstract. "The level at which discussions of 'breadth v. depth' and 'knowledge v. comprehension' are carried out and the 'structure' of the course is outlined."

2. the manifestational level which is concerned with "ways in which a student's achievement of an objective can be demonstrated".

3. the operational level which "defines an objective in terms of the precise means by which it is to be assessed", (114).

All of these authors have put forward arguments for the behavioural specification of objectives. Much criticism has been levelled at this stance. Eisner (46) makes a distinction between two types of objectives — instructional and expressive. Instructional objectives are those described earlier in this section and which specify what a student is to acquire.
after having engaged in a learning activity. Expressive objectives do not specify the outcomes, rather they describe an educational encounter. They identify a situation in which students are to work, a problem with which they are to cope, a task in which they are to engage: they provide "both the teacher and the student with an invitation to explore, defer, or focus on issues of peculiar interest or import to the inquirer." An expressive objective is evocative rather than prescriptive: it desires a diversity not a homogeneity of response. The evaluative task is not one of applying a common standard to the products produced, but one of "reflecting upon what has been produced in order to reveal its uniqueness and significance". In the field of science education, Atkin (3, 4) has been critical of behavioural objectives as they lead to the curtailment of writer-creativity, and for other reasons which are covered by Popham (105), a great advocate of behavioural objectives, who has collated a number of "reasons which educators employ to escape the practice of stating their objectives behaviourally". These have been the focus of much of the discussion of objectives. They are:

1. Trivial learner behaviours are the easiest to operationalise, hence the really important outcomes of education will be underemphasised.

2. Prespecification of explicit goals prevent the teacher from taking advantage of instructional opportunities unexpectedly occurring in the classroom.

3. Besides pupil behaviour changes, there are other types of educational outcomes which are important, such as changes in parental attitudes, the professional
staff, community values, etc.

4. Measurability implies behaviour which can be objectively, mechanistically measured, hence there must be something dehumanising about the approach.

5. It is somehow undemocratic to plan in advance precisely how the learner should behave after instruction.

6. That really isn't the way teaching is: teachers rarely specify their goals in terms of measurable learner behaviours; so let's set realistic expectations of teachers.

7. In certain subject areas, e.g. fine arts and the humanities, it is more difficult to identify measurable pupil behaviours.

8. While loose general statements of objectives may appear worthwhile to an outsider, if most educational goals were stated precisely, they would be revealed as generally innocuous.

9. Measurability implies accountability; teachers might be judged on their ability to produce results in learners rather than on the many bases now used as indices of competence.

10. It is far more difficult to generate such precise objectives than to talk about objectives in our customarily vague terms.
Discussion of Popham's defence of behavioural objectives has taken place, at length, in a monograph with Eisner, Sullivan and L. L. Tyler (106) and in detail by Macdonald-Ross (84). He summarises his critical review of behavioural objectives as follows:

"1. No consistent view exists as to the origin of objectives.

2. In the educational domain no well-defined prescriptions are available for deriving objectives.

3. Defining objectives before the event conflicts with voyages of exploration.

4. Advocates do not show how teachers can use objectives to guide unpredicted classroom events.

5. There are an extremely large number of paths through any body of knowledge, thus reducing the effectiveness of objectives in design.

6. In some disciplines criteria can only be applied after the event.

7. Objectives do not prescribe the validity of test items.

8. Objectives are inherently ambiguous.

9. The level of specificity problem has never been solved.

10. Objectives do not communicate intent unambiguously, especially to students."
11. Trivial objectives are the easiest to operationalise, and this is a problem.

12. The relevance of goal-referenced models of education can be questioned.

13. Weak prescriptions lead to cycling (through the curriculum development process). This can be costly.

14. Lists of behaviours do not adequately represent the structure of knowledge.

15. The use of behavioural objectives implies a poverty-stricken model of student-teacher interaction.

16. "The behavioural objectives scheme suffers from many of the weaknesses of any operationalist dogma."

He regards points 1, 2, 7, 9, 13-16 as the most important and 14, 15, 16 as the most fundamental. He concludes that behavioural objectives fulfil a useful role, but that they must be viewed with regard to their limitations. He sees the discussion of behavioural objectives as somewhat exhausted and suggests that attention is directed towards other problems.

Some empirical research which casts doubt on the ability (41) of anyone to write truly behavioural objectives has been presented by Deno and Jenkins. They asked teachers to rate a list of 45 verbs in terms of their observability. They calculated the means and variances on a five-point scale. Some verbs, for example, to solve and to recognise were rated low on observability, many of these were words
recommended (150) for their behavioural interpretation.

Much of the debate on behavioural objectives rests on the degree of behaviourality or precision that is required of objectives. The basic question is: How far can you take specification? Hudson (151) draws attention to the fact that behavioural scientists under strictly controlled laboratory conditions cannot specify in detail even the behaviour of an insect. So how much less likely are we to specify content of an educational experience?

Educational Objectives can be classified in different terms from those mentioned previously, the most popular being that of Bloom, Krathwohl et al (15, 71) commonly known as Bloom's taxonomy. Although originally proposed as an aid for the classification of test items, the cognitive domain has been used (15) extensively for the classification of curriculum objectives. The three domains of the taxonomy are:

Cognitive Domain covering knowledge, comprehension, application, analysis, synthesis and evaluation, (15).
Affective Domain covering receiving (attending), responding, valuing, organisation, and characterisation by a value or value complex, (71).
Psychomotor Domain covering perception, set, guided response, mechanism and complex overt response, (118).

This classification draws attention to areas of educational objectives which have been neglected in both curriculum design and evaluation. Stake and Denny (125) and Eiss and Harbeck (47) point out the lack of an affective component in assessment schemes. Kapfer (66), Williams (142) and Loree (81) propose ways of using the affective domain and
integrating it with the cognitive.

However, no matter how objectives are stated they still are the subject of much critical evaluation. Stenhouse, from the context of the English school situation, cites two practical problems in the stating of objectives and indeed any form of rational curriculum planning. They rest on two assumptions:

1. "that teachers who assent to lists of objectives agree in their values"

2. "that teachers who profess objectives will be able to operationalise them in the classroom"

He presents arguments against the soundness of both of these. His position is that it is an advantage to specify the content of a course, rather than objectives - "the content being so structured and infused with criteria that, given good teaching, student learnings can be treated as outcomes, rather than made the subject of prespecifications."(128)

The discussion on objectives has now turned full circle. No consensus has been reached on:

1. whether objectives should be specified
2. how they should be formulated
3. what function they should perform.

A case can be constructed from the literature to support whatever position it is wanted to adopt. It would seem that the differences of approach on the question of objectives rests on more fundamental values than those normally stated.
explicitly in the literature. To probe those values would depart from the main focus of this report. Perhaps it just reflects the different values of the diversity of people who operate in the educational system. If this is the case agreement would not appear to be possible.

2.32 Evaluation

Evaluation is a term which has many meanings and has been used in many different contexts for different purposes. Pace (97) states that "the term itself has lost almost all precision and perhaps much of its capacity to communicate among teachers, administrators, and researchers". The following sections will attempt to clarify and define the term evaluation in the sense that is being used in this report and discuss some of the issues which relate to the chosen concept of evaluation.

2.32.1 What is Evaluation?

Most authors who have written about evaluation have at some time tried to define it. The following are a few definitions in chronological order:

"the process of determining to what extent the educational objectives are actually being realised by the programme", Tyler (1949) (138).

"collecting and analysing whatever kinds of evidence are obtainable and pertinent to the matter under study", Sawin and Loree (1959) (112).

"the systematic attempt to gather evidence regarding changes in student behaviour that accompany planned educational experiences ", Harris (1963) (60).
"the collection and use of information to make decisions about an educational programme", Cronbach (1963) (40).


"a process for collecting and processing data pertaining to an educational programme, on the basis of which decisions can be made about that programme", Taylor and Maguire (1966) (134).

"the determination.... of the results.... attained by some activity.... designed to accomplish some valued goal or objective" Suchman, (1967) (132).

"the procedure by which programmes are studied to ascertain their effectiveness in the fulfilment of goals", Greenberg (1969) (58).

"the discovery of the nature and worth of something... ... to describe something and to indicate its perceived merits and shortcomings", Stake and Denny (1969) (125).

"the assembling and analysis of evidence prior to decision-making", MacKenzie, Eraut and Jones (1970) (85)

"to contribute to rational decision-making", Weiss (1972) (141).

The main elements in these definitions are that evaluation should:
1. contribute to decision-making
2. assemble and analyse data concerning the object of evaluation
3. make measurements to see if the objectives or goals have been achieved
4. assess the worth of the object.

The first two of these points are not contentious, however, points three and four are the subjects of discussion. Evaluation to measure solely intended student outcomes has been dealt with in an earlier section (see 2.222), the question of judgement in evaluation will be discussed here.

2.322 Judgement and Evaluation

The role of judgement in evaluation models has been mentioned in an earlier section. Some additions will be made here.

Wittrock (144) is clear about judgement in evaluation - "making explicit and measuring the bases of our judgements are central to the empirical study of evaluation of instruction". So too is Scriven (114). He defines evaluation as "a methodological activity which consists simply in the gathering and combining of performance data with a weighted set of goal scales to yield either comparative or numerical ratings, and in the justification of (a) the data-gathering instruments, (b) the weightings, and (c) the selection of goals".

Stake (123) in a paper on objectives, priorities and other judgement data puts this view forceably. He points out that all evaluation is based on data that is to some extent fallible, and that it is the responsibility of the evaluator to give evidence of the worth of the goals and objectives as seen by the different people in the educational process. He regards listing objectives as "selecting a few more-valued goals from
a vast multitude of possible goals". He concludes by pointing out that he has found few procedures which have been used successfully for making judgement data a part of evaluation.

Weiss (141) takes a different perspective on the problem and examines some issues in the role of the evaluator when making judgements. She suggests that it may not be possible for the evaluator to take such a threatening stance as to invite outside bodies to make judgements on programme activities. To do this would mean limiting some of the evaluator's other activities.

The role of judgement remains a standing issue in evaluation. Many authors advocate using judgement data but few practical examples of its use can be shown.

2.323 Types of Evaluation Activity

One basic distinction to be made is between formative and summative evaluation (114). Formative evaluation refers to those evaluative activities that take place during the formation of a programme. It is primarily concerned with gathering data that can be fed into the programme, aiming towards its subsequent improvement. Summative evaluation takes place on a finished or nearly finished product, curriculum etc. It aims to produce a profile of its strengths and weaknesses - to sum up the programme. It may involve the comparison of one course with another.

Formative evaluation is similar to what Cronbach (40) calls 'evaluation for course improvement'. He declares that "the greatest service evaluation can perform is to identify aspects of the course where revision is desirable". He thinks that the comparison of one course with another should not
dominate evaluation plans.

Another distinction has also been made by Scriven (114) and discussed by Eraut (50). That is between intrinsic and payoff evaluation. Eraut calls payoff evaluation, performance evaluation. Intrinsic evaluation is essentially an armchair activity concerned with the analysis of an existing or proposed curriculum in order to discern its likely planned and unplanned effects. Performance evaluation is concerned with finding out to what extent these effects are realised in practice. It includes both an assessment of the extent to which intended objectives have been achieved and an attempt to detect whether there have been any unintended outcomes.

The work reported later in the report will be chiefly of a formative, performance nature, so that attention will be focussed on these areas.

Evaluation and Hypothesis Testing

Some authors (112, 128) have accepted the criticisms of the limitations of using an entirely objectives-based evaluation strategy and have put forward an alternative based on hypothesis testing.

Sawin and Loree base their work on a Tylerian evaluation model but recognise some of its deficiencies with respect to unplanned effects. They ask the question, "What is to take the place of objectives when we evaluate unplanned effects of an educational program?". Their answer is to turn to the core values of the programme and concepts that derive from them and from these to set up
hypotheses about possible outcomes, intended and unintended. They suggest the following criteria for selecting hypotheses:

1. The importance to the objectives, the general purposes of the school, and the core values of those in control of the school.

2. The requirements for an adequate sampling of student behaviour patterns that are related in an important way to the instructional programme in question.

3. The feasibility of testing the hypotheses with the instruments and procedures available.

4. The willingness of the teacher and other school personnel to have the hypothesis testing - that is, the presence or the absence of threat in the hypotheses.

They claim that when this modified approach to (Tylerian) evaluation is used, "the evaluation process can be started at a point closer to the phase of the cycle where useful feedback information is available.

Stenhouse (1969) from his rather different attack on the limitations on the use of objectives (see section 2.212) also puts forward the idea of hypothesis testing in evaluation. He suggests that "either from past experience, or from exploratory studies or from theory ....... hypotheses may be generated regarding the possible range of effects of a given curriculum specification and their variation in relation to the web of contextual variables in schools". From these
hypotheses, he says, some would be selected as crucial and tested.

Hypothesis testing would seem an attractive alternative then to the long process of writing objectives. However, neither of the authors give any guidelines as to how one can go about generating and selecting hypotheses in detail and to what areas of study they should be related.

2.325 Non-prescriptive Evaluations

It was mentioned briefly in 2.21 that the presently accepted position in evaluation was that it was not possible to prescribe evaluation strategies without a detailed knowledge of the goals of a given programme. This position is currently being challenged from two points of view.

Firstly, Scriven (115, 116) proposed a different view of evaluation. He has become increasingly uneasy about the separation of goals and side-effects and puts forward "that consideration and evaluation of goals was an unnecessary, but also a possible contaminating step". He worked on an alternative approach - "the evaluation of actual effects against (typically) a profile of demonstrated needs in this region of education". This he called Goal-Free Evaluation.

It is put forward as an addition to internal formative evaluation, rather than as a substitute. "The less the external evaluator hears about the goals of the project, the less tunnel-vision will develop, the more attention will be paid to looking for actual effects (rather than checking on alleged effects)."
Commenting on Goal-Free Evaluation, Stufflebeam (131) says that this is one methodological strategy that can be used to supplement others including goal-based evaluation and the evaluation of goals, that it should be regarded as an additional tool in the evaluation repertoire.

Secondly, Parlett (99 ) makes a criticism of goal-based evaluation strategies and of any approach which prespecifies the goals of evaluation. He points out that in any innovative programme it is impossible to assess the most important questions to be asked of an evaluation before the start of the study. His argument is based partly upon the practical criticisms levelled at any detailed specification of objectives ( 3 ), but in addition he proposes a strategy which he calls the 'social anthropology paradigm' as an alternative to the 'agricultural botany paradigm' which he claims to underlie much evaluation research. This paradigm, developed further in collaboration with Hamilton (100), is closely related to the methodology of 'psychiatrists, management consultants, sociologists basing their research on ethnomethodology (53 ) or on symbolic interactionism ( 17 , 88 ) and indeed teachers themselves' (99 ). He calls this form of evaluation illuminative evaluation. ( 13)

This approach is characterised by two basic principles ( 9 ). They are that the research priorities should not be determined solely by the investigators' interests, but also by those people in the situation that is being studied, and by others in similar contexts; and that the course to be evaluated must be studied in its total context, within the social and intellectual environment in which it is found. That is, the problem areas to be investigated in an evaluation should not be specified exclusively in advance, and the evaluator should
concern himself with the learning milieu, of which the courses to be studied form a part.

It has been proposed as an addition to the repertoire of evaluation methods to be used in the initial and mid-stages of an investigation (59) and as one of a wide range of different strategies that should be directed at any given problem (91).

Illuminative evaluation differs from goal-free evaluation proposed by Scriven in that no attempt is made to ignore the goals of the people who are responsible for the programme that is being examined. It takes an eclectic approach and uses all the information that can be obtained from as many sources as possible and uses that to illuminate important issues.

2.4 The British Context

The majority of the studies reported in the previous sections related to work done in the United States and work in the school system. This section will briefly sketch some of the contributions to this area in this country with particular regard to higher education.

2.41 Course Evaluation in Higher Education

Beard (7) reports on the paucity of studies on any form of evaluation in colleges and universities in Great Britain. Most of the attention that has been given has been on the teacher, studies of comparative teaching methods (14) and assessment. Indeed, Beard and Bligh's (8) comprehensive review of research into teaching and learning in higher education is entitled "Research into Teaching
Methods in Higher Education", and the space devoted to evaluation of courses in the sense used in this present report is less than one page out of seventy-seven. Similarly, in her Penguin (7), the chapter on "Evaluation of Learning and Teaching" is devoted to the discussion of examinations and the assessment of lecturers. Also in Butcher and Rudd's (29) substantial compilation of papers on "Contemporary Problems in Higher Education" the chapter by Parlett, the only one on evaluation, mentions no examples of British evaluation exercises.

Some work has been done in a context closer to that of Britain than the United States, in Australia. Falk and Dow's monograph (51) presents examples of work they have done in collaboration with courses at the University of Melbourne which can be classified as formative, performance evaluation. At the University of Sussex, the Centre for Educational Technology has been studying some of the problems associated with the role of evaluation in higher education (85), and Eraut (50) has reported work on intrinsic and performance evaluation of material on the Inter Universities Biology Teaching Project. On the same project, Dowdeswell (43) expresses the view that there is a lack of trained evaluators for working on curriculum projects, which is not surprising if no evaluation work is taking place.

One major exception is the Open University. A procedure for formative evaluation of course materials known as developmental testing has been in operation since the university was established (93) and evaluation measures have been integrated into other parts of their systematic course development procedures (78).
2.5 Summary

This chapter has presented some of the research on evaluation which has been done elsewhere, some practical, but much of it theoretical. It is this which will be used as a conceptual framework for the practical evaluation exercises reported later. Little consensus is available on any questions concerning evaluation; one point of view can always be countered by its opposite. Within any single approach agreement on procedures begins to be reached, but the present state of affairs is probably truthfully represented by Bruner, (28) quoted in Whitfield and Kerr (146).

"Such is the latitude in the choice of criteria for evaluation, that something nice can usually be said about any course or curriculum!"
3. SOME ISSUES IN EVALUATION RELATIONSHIPS

3.1 Introduction

The previous chapter gave a view of course evaluation as found in the literature of educational evaluation. However, this only provides a part of the total perspective in which course evaluation takes place. Other considerations are infrequently reported in the literature, but they are of equal or greater importance to those that are reported (31, 141). The factors concern the contexts in which formative evaluations are based and the constraints on evaluation activities that are related to these, and the roles and relationships that evaluators adopt and work with in these contexts.

The aim of this chapter is to explore some of the issues involved in the contexts that evaluations are based in and the relationships that can be adopted by an evaluator. This leads to two other important issues in evaluation which need to be considered. Firstly, there is a problem about the kind of information that constitutes evidence in evaluation and the criteria that should be used to judge the worth of that evidence. These will be discussed and a set of guidelines for judging evaluations developed by Stufflebeam et al (130) will be described.

Secondly, in anticipation of Part II, the chapter will conclude with a section about the biasses of the evaluator. In any evaluation the evaluator's interests, experiences and values will influence both the design and execution of a study. Sometimes this, often unconscious, bias can be obscured by an apparently rigourous evaluation design, but the
evaluation design can be thought of as the explicit bias of the evaluator or the tradition in which he works. In evaluation which does not follow a rigorous design, evaluator bias will, in addition, affect many of the interactions which comprise the investigation. In recognition of this a section describing the recognised bias of the present author has been included.

3.2 The Evaluation Activity

The evaluations centred on the problem of obtaining information from students about their courses that could subsequently and directly be used as a basis for course improvement. The specific problems related to this were:

1. What kind of information is useful for this purpose?
2. How can it be collected?
3. What is the researcher's role in this situation and what does he do?

A detailed discussion of the first two points will be included in the Chapters related to each project, but it is appropriate for some general considerations to be made here.

1. The type of information required is dependent upon the use to which it is to be put. Different information would be gathered if the prime emphasis was on grading students than if the emphasis was towards course effectiveness. In the first case data on student performance would be most relevant, whereas a much wider range of instruments is needed to measure course effectiveness.
2. The type of information required would normally depend on who was going to use it. Was it for the course organiser to plan more appropriate teaching strategies, or the head of department to assess a lecturer's competence?

3. The type of information required can be dependent on the amount of resources that are available to support changes. If it is totally unfeasible to radically restructure a course should information pertaining to this be gathered?

4. The type of information required can depend on the particular interests of the lecturer concerned. If he has strong ideas about problems in certain areas of the course to what extent is it appropriate to consider only these? This is a very real problem is research resources are very scarce.

These are a few of the immediate problems concerning the research activity. Others are related to the constraints of the situation which form the next section and the role of the evaluator following that. However, it is necessary at an early stage to describe in detail the common elements of the three case studies.

3.21 Common Elements

For this piece of research the researcher adopted the role of an educational technologist primarily interested in course evaluation (78). He was based in a university Institute for Educational Technology and worked in conjunction with teaching staff in departments of the same university. He was introduced to these staff as a research student working on a project which aimed to evaluate some selected university science
courses. Their co-operation was sought to assist him in this task by allowing him to study one of the courses that they were responsible for teaching. The Quantum Mechanics course was rather different as it was given by the research student's supervisor (see Chapter 5).

The situation was, then, that lecturers were working with a research student interested in evaluating one of their courses. In most cases the student was relatively unfamiliar with the subject matter of the course. The lecturers were committed to this situation only in so far as their interests and time would allow, no external incentive was provided.

This can be characterised by the situation of two people – the evaluator, or researcher, and the subject expert, or teacher, – working together on the joint problem of how to evaluate a given course; both parties are in the relationship through choice, without formal external constraints, in a research orientated atmosphere.

3.3 Constraints

The situation, however, is not as straightforward and clear as it might seem from a description of the apparent relationship. Many factors influence the two people in this situation. Some are very specific and apply only to the particular course and people concerned, others are general for the institution or setting of the work, and yet others apply to any kind of collaborative research. In one sense every piece of work in this field is unique in that it deals with one course in one subject with one member of staff. In another situation, all the variables change, but also so
do the more subtle environmental effects. It is these that this section attempts to document.

3.31 General Constraints

Some factors affect all collaborative research efforts. Those relating directly to the type of involvement here are:

1. Commonality of aims, i.e. do both people have the same aims for the collaboration?

   Do both parties enter with the same expectations? They have different interests; one has a subject-matter/teaching orientation; the other a research/investigative orientation. One is concerned with immediately useable, short-term results; the other with longer term, more methodological issues. Their aims may not be expressed in the same language. The evaluator has available the precise, analytic language of the literature; the teacher has his everyday usage. This problem is compounded by the tendency to use everyday words, eg. objectives, in precise ways.

   They may agree on the same statement of the aims of the research, but do they interpret them in the same way?

2. Confidentiality

   It is a norm of undergraduate teaching that no-one sees a teacher's performance except the students. By engaging in research on teaching one is imposing a possibility of exposure to a wider audience, particularly to other staff in the institution. Teachers, like other people, often prefer their less.brilliant activities to be quietly forgotten. Norms of confidentiality can be, and should be, established by the evaluator, but the collection of data, even with an assurance of complete confidentiality, can result in a limitation on the range of possible activities that can be attempted, and
to be rather less significant than the actual changes.

Another implication of this strategy was that students were likely to perceive evaluation as an optional extra and so take it less seriously. This may be especially true as all questionnaires and tests were explicitly non-compulsory. This does produce the danger of a biased sample. It can be minimised by adequate distribution and collection facilities, but there is inevitably a tendency for the weaker students, and those not regularly attending the course, to be under-represented in the sample. The actual effects on the results are not known, but knowledge that the sample is so biased must be taken into account when interpretations are made.

The evaluation approach also aimed to be non-threatening and minimally intrusive. Steps were taken to assure students that the results of tests and questionnaires for evaluation purposes would not be used to assess them. It was frequently emphasised that names need not be put on questionnaires or, if they were, the names would not be revealed to the academic staff. In the case of responses to open-ended questions, student replies were transcribed before being presented to staff if there was any likelihood of handwriting being recognised. Questions were asked that were associated with students' experience of the course: background variables, such as previous examination grades, were not probed.

Secondly, the evaluator adopted a teacher-centred role. In investigating the teaching and learning situation, the evaluator can be seen to have two clients - the teacher and the students. The needs of these two groups can only rarely be met simultaneously in the present system of university courses. In this case a teacher-centred role was adopted
3.32 Situation Specific Constraints

The situation we are dealing with is that of university teaching of science subjects in introductory courses.

Factors which relate to this are:

1. The examination system.
   A present component of the teaching-learning system in universities is the assessment of students. Assessment takes place for a variety of reasons, one of the most important being to grade students, that is, to assign worth to individual students for various given standards of performance in examinations. This component acts as a constraint on course evaluation because of the highly prescribed nature of the forms of assessment. Usually the form of examination is given, it is imposed by departmental or institutional requirements. It is not necessarily open to change as part of evaluation research. This was so for all the courses studied.

   Student performance can form a very useful measure in course evaluation. However, at present examinations are not designed for this purpose, and indeed, the 3-hour essay-type examination is very ill-suited. Detailed objectives for questions are not normally specified, the criteria for marking is obscure and the range of possible objectives covered is very small.

2. Organisation.

   Amongst the organisational factors can be included the time-table and the time constraints of the examination system. The time-table specifies when and where staff and students are to present themselves. Times for meeting outside these hours
are limited, especially in science courses, and are dependent on a high degree of enthusiasm by both parties. Courses are arranged, typically, in 5 or 10 hour units to fit into terms and to fall before exams, and a syllabus is provided which should be covered. Very little time is available for non-syllabus orientated activities and this includes course evaluation.

3. Teachers.

Engaging in systematic course evaluation is not one of the normal activities of university teachers. Their courses were not originally designed to take evaluation into account. They are unused to time spent in detailed discussion of their course with an outsider, especially a non-expert in the subject. Staff are orientated to the norms of the examination system and arrange evaluation as an appendage, if time permits.

4. Students.

Although the aim of course evaluation is ultimately to improve the course, the time scale may be too long for the students. They may perceive and express defects in the course which they expect to be changed immediately. If they do not see change taking place, especially if it is promised by the evaluator, then they are likely to be resistant to co-operation at other times in the course.

Another constraining effect of students is rather more fundamental. It may be that the aspects of the course that the lecturer sees as important for evaluation are not similar to those that the students feel important. If students perceive evaluation to be irrelevant and useless to them they are likely to reject the idea, either directly by
expressing this or indirectly by opting out of filling in questionnaires, etc. Student apathy, although it can be a significant indicator in itself, does not assist the unambiguous diagnosis of course problems.

What underlies many of these constraints is inherent in any context. It is that evaluation is always a second-order activity. The important object is what is being evaluated and not the evaluation itself. Priority must always be given for resources for the course that is studied. The lecturer must spend most of his time on the subject of his course. The students' attention must be given to the substance of the course. Evaluation may be a vital component to a course, but it will always remain an optional extra unless external pressures make it otherwise.

3.4 The Role of the Evaluator

This section describes the role that was finally adopted by the researcher in the present study in the light of the research problem and the constraints imposed on it.

Firstly, the evaluation strategy started from leaving the course as it was and not making any organisational changes to adapt it for convenient evaluation. The evaluation was then an appendage to the course and not an integral part of it. This presents severe limitations and in some cases makes the evaluation design less rigorous. For example, in order not to be too disruptive, the before and after measures discussed in Chapter 6 were not strictly 'before the students had met the lecturer' and 'after all lectures had finished'. This particular restriction would add a conservative factor to any differences found causing, probably, the changes measured
for three reasons:

1. It is easier to work with one or a few people than it is to work with many, especially if there is likely to be a substantial difference in perspective.

2. The teacher controls the course in so far as he presents certain experiences to students and is responsible for minimally preparing students for examinations.

3. Teachers are more readily accessible to discuss and analyse courses they are primarily responsible for.

Given that little evaluation research of any kind has been undertaken with university courses, either teacher-centred or student-centred, it seemed sensible to select the present teacher-centred as the most convenient.

In this context, what is meant by a teacher-centred role is that the researcher attempts to evaluate the effectiveness of a course from the viewpoint of the value structure of the teacher. That is, the factors in a course that are highly valued by the teacher will be given the greatest attention by the researcher. However, this does not mean that information about the course outside this framework will be ignored. The evaluator is charged with the duty of being open to the expected and the unexpected, the valued and the apparently trifling.

Such a role involves close association with the teacher. In this respect the researcher should aim to be non-directive
non-threatening, non-judgemental and have a detailed understanding of the teacher's perception. This, in practice, will often involve him in the situation of acting as a counsellor on teaching problems as well as that of investigator of cognitive content.

Obviously, this is an ideal, and some deviations from this are inevitable. One example of this can be seen in the present study where similar methods were used in two different courses in two different contexts - the aims questionnaire. These were obviously based on a strategy by the evaluator that lay beyond the total teacher-centred perspective described above.

3.5 Evaluator and Teacher

If an exclusively teacher-centred role is adopted by the evaluator then it is necessary to determine that the activities of the evaluator could not equally well be carried out by the teacher. If the teacher could evaluate his own course then the evaluator would become redundant and a saving would be made. This is a criticism often voiced about evaluation by university teachers. It is usually put in the form, 'well, what you are doing is just good teaching, and the best teachers do it anyway'.

It is not intended to defend the expertise of the evaluator by mystifying his role - it is a role which, it is likely, could be filled by as many people who could become teachers. However, there are certain characteristics of the teacher's role and the evaluator's role in a particular context which make it very difficult, if not impossible, for them to overcome the differences and exchange roles.
The teacher is primarily responsible for enabling the students to learn. He is the subject-matter expert in the teaching situation. He is the designer of the learning experiences that he presents the students with. It is necessary for the success of this task that he becomes committed to, and enthusiastic about, what he is doing, and is able to transmit this to the students. In practice, the teacher personally identifies with the particular teaching strategy that he is adopting and views teaching from the perspective which he has constructed.

The evaluator is also committed to the improvement of teaching, but he does not necessarily have the same perspective as the teacher. The evaluator is responsible for helping to determine the teacher's specific aims and strategy for the course from the feedback to the teacher of reliable observations of the course and for the transmission to the teacher of information about how students are reacting to the course. He is able to tackle this task more easily no matter how negative the student comments are, because he is not the subject of their views. He is more able to obtain reliable information because he is outside the immediate teacher-learner interaction. The evaluator does not, or should not, have a personal investment in the particular course or teaching strategy that he is working with. He can direct his full attention to eliciting relevant and detailed feedback, both positive and negative, from students and to presenting this information in a constructive, rather than destructive, way to the teacher.

The incompatability of the two roles is not because of any technical functions that can be carried out solely by the
teacher or the evaluator, but rather because the two roles require a different commitment. Not all teachers are like William James (see frontispiece) in their capacity to combine the two perspectives.

This conclusion is not supported or refuted by published evidence as far as I am aware. However, some personal observations of the reactions of teachers to the use of lecture feedback questionnaires at two other universities may be indicative. In some cases the teachers received feedback which was critical of their lecture course. They tried to make changes to the course in order that in the following year a better course could be run. They found that almost identical student comments were made in the following year. The teachers told me that although they appreciated what the students were saying, they were unable to relate this to their actual teaching, in other words, they could not link the students' observations to their own performance. In one case, the teacher continued to use the lecture feedback forms, but he said he did not try to change in the light of it.

It may be that this was an isolated case and that the teachers I met were especially inept. I do not believe this to be the case. Even if it were, it would still point to the conclusion that some teachers, at least, are not able to take on the role of both teacher and evaluator.

This discussion should not be taken as an argument against teachers' monitoring their own courses. Self-monitoring is a necessary and vital function for any teacher and should be one of the basic teaching skills. However, at
present, some university teachers manifestly do not have the appropriate skills in self-monitoring with respect to their teaching activities. The lack of sensitivity about how lecture courses are being received by students is but one indicator of this. In situations where self-monitoring is not fully developed then an external supportive evaluator can fulfil an important role in the improvement of teaching. It is not necessary that every teacher's course be the subject of intensive evaluation: that is totally unrealistic. It is not unrealistic, however, for teachers to consult with evaluators who, adopting an appropriate role, can help a teacher approach the situation where monitoring of teachers ceases to be externally provided and becomes an integral part of the course.

3.6 Evidence in Evaluation

Once it is accepted that the difference between the role of evaluator and of teacher is not purely one of different function and technical competencies, then the notion of what is acceptable evidence in evaluation becomes quite complex. It is no longer possible to regard the gathering of information as a value-free, objective enterprise. If it is necessary that the perspective of the evaluator should be based on the teacher's construction of the situation that is being evaluated but not subordinate to that, then the kind of evidence that it is meaningful to collect and the interpretation of that evidence become problematic.

It is obviously necessary that if some measurement of a course is being employed then it must be as reliable as possible and it must be a valid measure of the activity or object that it is intended it should measure. This kind of
evidence, which can be called scientific evidence of the worth of an evaluative activity, is incomplete. It is the kind that is acceptable in scientific research, but it is not a complete set of criteria for evaluation.

Stufflebeam and his colleagues (130) have examined the question of what are the appropriate criteria in evaluation for decision-making. They describe three types of criteria: scientific criteria, practical criteria and prudential criteria. The first of these are the criteria that would normally be applied to the scientific investigation of an activity and include validity, reliability and objectivity. Practical criteria can also be applied to scientific problems, but the lack of them does not necessarily devalue the scientific nature of the task. These relate to relevance, importance, scope, credibility, timeliness and pervasiveness. It is arguable as to whether Stufflebeam's final category which includes a single prudential criterion warrants separate status. His criterion of efficiency is very similar to the concept of Occam's razor in science.

In detail, the criteria are:

Scientific Criteria

(a) Internal validity
There must be a close, if not a one-to-one, correspondence between the information that is being obtained and the phenomena it represents.

(b) External validity
Does the information hold only for the sample from which it was collected or for other groups (or the same group at other times) as well? Is the information generalisable?
(c) Reliability

Is the information replicable? If new data is gathered would the same findings result?

(d) Objectivity

Would different observers from different viewpoints make the same reports? Would different judges deduce the same event?

Practical Criteria

Scientific criteria can be judged independently of the receiver of the information. In science, the ultimate receiver is the scientific community. In evaluation, the needs of the receiver, or client, must be more carefully considered.

(a) Relevance

Does the evidence obtained relate directly to the purposes of the evaluation as defined by the sponsors? This is, in effect, a more precise form of validity.

(b) Importance

What measures does the client judge to be the most important? What does he give high significance to?

(c) Scope

Does the evaluation cover all the important and relevant areas? Is it restricted in any sense?

(d) Credibility

Is the evidence to be believed by the clients, even if it fulfills all the other criteria?
(e) Timeliness

Is the evaluation reported in time for decisions to be made in the light of the findings?

(f) Pervasiveness

Do all the people who need the information have it available to them? Is it effectively disseminated?

Prudential Criterion

(a) Efficiency

Once all the other criteria have been satisfied, is the evaluation chosen the most efficient one possible in terms of time, money and resources? Is it the simplest?

Stufflebeam and his colleagues state that they believe that few of these criteria are used for most evaluations that they have examined. It may be that they form an unattainable ideal in practice and that they should be used as a set of cautions for practitioners and guidelines for critics of evaluation studies. The number of social science studies reported in the literature which fulfil all of the scientific criteria alone and still contribute something worthwhile is very small. With the addition of seven additional factors the chances of a single piece of work meeting all of them appears to be even smaller.

If this pessimistic assumption is substantiated then the most important criteria should be examined and these should be given the most weight in practical evaluation.

Of the eleven Stufflebeam criteria, two can be considered as first order factors - relevance and importance. Without
these, all the others are unnecessary. If the purpose of the evaluation is not pursued by the instruments and and the evidence gained is not valued by the client of the evaluation, then an examination, of the scientific criteria in particular, is redundant.

These two criteria are also those most related to the teacher's perspective. Indeed, Beard (6) points out that university scientists and engineers do not de facto use the same criteria for their teaching as they do for their research - the scientific criteria.

In small scale evaluation, as in the present case, it becomes important to examine priorities because only a few activities out of the many desirable are practicable to pursue. The rest have to be approximated by substitutes which are often much inferior to those ideally desirable.

The Evaluator as an Instrument

It is possible in many circumstances in course evaluation to use measures which have proved useful elsewhere and which have demonstrable reliability and validity in given circumstances. However, in small scale evaluation it is often not feasible to use more than one of such methods because of the demands it makes on an evaluator's time. For example, use of Flander's interaction analysis or a reportary grid method is very powerful in many situations but it is very demanding. A decision has to be made on whether collection of information in a narrow context has higher priority than the use of less systematic methods used on a broader front.
In some situations, the use of less systematic methods will be inevitable on these grounds alone. However, there are much more positive grounds for the use of non-standard measures. In particular, the use of the evaluator himself as an observer will be discussed.

Of all the data gathering instruments in the evaluation context the most powerful and potentially the most sensitive is the evaluator himself. In as much as all the evaluation instruments are derived from him, he is the main influence on the total evaluation procedure. In most situations, the evaluator will interpret the situation he is due to investigate and from his informal interpretation formalised, external measuring devices will be constructed. In most social science contexts only the results derived from the finalised instruments, such as questionnaires and interview schedules, will be presented. However, the evaluation setting differs greatly from the research context. The most important difference is in the quality of data that is used to make decisions. In the research setting only data of the highest quality should be used in making interpretations and drawing conclusions. The evaluation situation is different. Decisions will be made about a course, no matter how good or bad the quality of the evidence is about a course; they will even be made if formal evidence is totally absent.

In evaluation, all the evidence that can be gathered has the potential of being used to influence decisions. If the odds of making an appropriate decision about a course are increased to anything more than 50-50 then the evidence is worthwhile. This is not to excuse the collection of inferior information, but to suggest that in many situations second class data is better than no data at all, or even, that
second class evidence pertaining to the decisions to be made is superior to first class information about a problem outside the realm of immediate discourse.

While evaluation is not research, there is a need for research into evaluation. This is a complex activity, an order of magnitude more involved than evaluation studies themselves. In the examination of some fundamental issues in evaluation this thesis aims to help clarify the problem area to which research into evaluation should be directed.

Falk and Dow (51) quote Hemphill (61) on this point: "Confidence in a conclusion, as represented by the research convention implied by the general acceptance of the '0.05 or 0.01 probability level' as the criteria for 'belief' of a research finding, is a luxury a decision-maker seldom can afford. Rather more frequently he faces situations where any information more dependable than that provided by a 'flip of the coin' is desperately needed."

Evaluation demands that resources should not be wasted, cf. Stufflebeam's efficiency criterion. The observations of the evaluator, even when they are not systematised, must under many circumstances be recognised as valid information in a particular situation. An aim of good evaluation should always be to make these observations and interpretations explicit and open to replication, but while this is happening the course being studied will change and the evaluation must keep in touch with it.
It is inevitable that in a highly applied, project-orientated context, working with real problems outside a laboratory, some degree of subjectivity and lack of control of potentially important variables will occur.

An (apparently) objective account of the three projects could be presented. It would be inappropriate though, since it would neglect essentially those factors which affect their reproducibility. The evaluation studies were done as a part of a dynamic process. It would not be possible to control the different factors without controlling the process itself. It is not the intention of these evaluations to do that. If a traditional lecture course is being investigated it must remain a traditional lecture course otherwise one is investigating something different. There is, of course, no escape from this. It is inevitable that the measuring instruments will to some extent affect the activity that is being measured particularly in the teacher-centred approach adopted. If this is bound to happen, then it is desirable that its influence is a positive one. Again, there is no way of guaranteeing this, but, it should be a direction in which to aim. Some of the positive outcomes of this evaluation, as well as giving information on areas for course improvement, would be increased teacher awareness of his teaching activities and increased student awareness of his learning activities.

One of the directly unmeasurable influences in this process is the bias of the evaluator. This can affect not only the interpretation of results but also the choice of evaluation strategy. It is relatively straightforward for
results to be re-analysed and re-interpreted; it is not possible for the research to be repeated using a different strategy without another complete project. It is important to be able to assess the degree of influence of the evaluator. However, there is no procedure for this in the present circumstances.

Some of the factors which may be relevant to the evaluator's approach in this situation have been mentioned already, i.e. an interest in the process of evaluation as well as the products and an emphasis on evaluation methodologies and strategies. Others are: a commitment to a position of the importance of active student learning in the teaching process; the importance of different perceptions of what is being taught; an interest in innovative teaching methods; an acceptance of the vital role of students in the improvement of teaching; a dislike and distrust of present examination methods; and an acceptance of the possibility of radical change in the teaching and learning situation as being the minimum change necessary for significant improvement in learning.

Many other factors may well be relevant, such as age and previous educational experience. Those factors regarded by the evaluator as having importance have been stated.

These can only cover some of the bias effects. An alternative strategy to elicit the effects more rigorous would be for representatives of groups holding different value positions to contribute statements of perceived evaluator bias. This approach is worth a project of its own and it cannot be adopted in the present study as it is not likely that such diverse groups as those required could
adequately be selected by the present evaluator from his own value position.

3.8 Summary

This chapter has aimed to draw attention to the factors important in the practical context of evaluation research, particularly in a university. Some factors have been stated, but it has not been possible for systematic empirical evidence to be assembled to support their statement. They are regarded by the author as having sufficient importance to warrant their inclusion on these grounds alone. Most of them cannot be controlled or altered by an evaluator: to do so would be to change the research. For this reason it is essential that evaluation researchers should develop ways of monitoring and documenting them. Only in this way will the importance of evaluation evidence be able to be judged by the people who have initiated it.
PART II THE CASE STUDIES

Chapter Four: An Introductory Physics Laboratory Course

Chapter Five: An Introductory Biology Lecture Course

Chapter Six: Quantum Mechanics and a New Teaching Method
Part II presents the three case studies, which form the core of the thesis and represent the empirical evaluation on which the discussions of evaluation are based.

Each study has an introduction which briefly describes the context of the course that was examined. Details of the structure and organization of the course are given and some of the constraints that were faced by the evaluator are described. These constraints are those pertaining to the particular course or type of course rather than the general constraints of the university context mentioned in chapter three. A section has been included entitled 'Rationale for Choice'. In this an attempt is made to determine why the particular course was chosen for evaluation. The choice of course relates closely to the constraints of the situation and it can be thought of as a description of the initial context which faced the evaluator.

Some of the studies include a section on previous work related to the teaching problem or teaching method where this is directly relevant. The evaluation approach adopted is presented with the procedures, results and analyses. The immediate outcomes of the studies are discussed and each chapter concludes with a reflective assessment of the final outcomes and limitations of the evaluation.

There is an inextricable relationship between general problems found in evaluations and in particular problems and issues in the studies reported in the following chapters. The chapters are presented in a style which relates the two throughout. In many ways it represents the dialogue that an evaluator has with himself and his colleagues when deciding on a particular strategy in a given situation.
4. CASE STUDY I: AN INTRODUCTORY PHYSICS LABORATORY

4.1 Introduction

This chapter gives an example of a course evaluation which was developed to provide information about a first year physics laboratory course that could be used to provide a basis for course improvement.

In the course which was examined certain innovations in teaching methods had been introduced. These included the use of tape/slide/film presentations (49), and self-service programmed experiments (24). There was a commitment to innovation on the part of one group of staff who were responsible for the course, and this group was also committed to evaluation, but less strongly.

The problem which was originally formulated in this case was to devise a method of evaluation that could be used to examine the innovative and the traditional components of the course from the same viewpoint. This could have introduced a great difficulty: if the innovative and traditional components of the course aimed at achieving disparate goals then viewing them from the same position might lead to a misrepresentation of one of the types of experiments. One group of experiments might be judged as failing in terms of goals that the designers of these experiments had rejected. However, this situation did not arise as staff agreed that it was necessary that the new experiments succeeded in achieving new goals and that they had also to be shown to be worthwhile from the perspective of the established aims of laboratory work.

The initial problem was set: what are the established aims of laboratory work in physics? The answer to this is not obvious. Although some work had been done on the aims of physics laboratory work in universities in the U.K. (34), the aims developed, and the results produced by that study could not be presumed to be directly transferable to the evaluation
situation of the course at hand. Thus some time was devoted to a consideration of the teaching aims in first year physics laboratories.

4.2 The Course

The course was studied in 1969 and 1970. It was a first term, first year laboratory course in physics provided for students studying physics, physical science, electrical engineering and metallurgy. It was part of a common two term course for these students organized jointly by four departments. Each department provided a laboratory course in its own subject area which was taken by students in all departments. The overall course aim for these activities was to show students something of the kind of laboratory activity appropriate to each subject. The physics laboratory contained, for example, physics experiments supposedly typical of the kinds of subject matter and kinds of activity that physics students would study in later courses. Each laboratory course took four sessions of four to five hours each during one term.

The physics laboratory was composed of the following: two experiments each taking one session to complete, one experiment of one session using a programmed script involving tape/slide and film sections, and finally a group of four to five experiments comprising one session. The first two were known as traditional experiments as they conform to the norm for experiments in physics laboratories. That is, they have a script giving instructions on what students should do and things they should investigate without specifying performance levels, and often a collection of apparatus that has to be assembled by the student is provided. A demonstrator is available to discuss problems and generally help the student complete the given instructions. The experiment using tape/slide/film was not included in the evaluation as it was the subject of an independent investigation (63): it will not be discussed further. The final group of experiments were known as self-service experiments as they were designed to be
performed without the presence of a demonstrator. They had specific behavioural objectives, programmed scripts, pre-assembled apparatus, and were each designed to take about one hour to complete.

4.3 Specific Constraints

In addition to some of the general constraints on evaluation activities mentioned in Chapter 3, there are some affecting laboratory courses in general and this course in particular.

Firstly, it is generally recognized (see 4.5) that in laboratory situations it is difficult to assess student performance. It requires a large investment of time and resources and the demand on students may be of the same order of magnitude as the course itself. So simple measures of student performance are difficult, if not impossible, to obtain. As there was a tight schedule of activities planned for the course additional time was not available for this purpose and it was judged extremely unlikely that if such time was requested it would be given.

Secondly, it is common in a laboratory course that, unlike a lecture course, many staff are responsible for running the course. It cannot be assumed that all staff have identical aims for the course. In this case the number involved was seven. These were drawn from two departments and they fell into two groups. One was committed to innovations and they were responsible for introducing new types of experiments into the course; the other was composed of the staff who were formally responsible for the course, they had been concerned with the design and administration of the course prior to the introduction of the new experiments.

Finally, there was a constraint due to a particular role conflict on the part of the evaluator. The evaluator was identified with the group of innovators and he had in the
previous year (1969) been involved in the design of some of the self-service experiments. From the perspective of some of the staff he may have been regarded as having a prior commitment to a particular outcome of evaluation, that is, the validation of the self-service experiments. However, the evaluator was not conscious of this interpretation at the time, or of the possible role conflict that may have existed.

4.4 Rationale for the selection of the course for evaluation

Why was this course chosen to be evaluated? If it was a special case, are the findings in any way transferable to other courses? There are many reasons why a particular course is selected for attention. These are both formal and informal. It is necessary to examine them closely in order to check that the conditions under which the evaluation took place were not so unique that no conclusion can be drawn about the evaluation of any other course.

The initial reasons for choice of the course do not stand unaltered by time and by the interaction of the evaluator with the staff and students in the course. The reasons the evaluator continues with his investigations are not necessarily the same as those that encouraged him to start work on the evaluation.

The prime reason for choosing the course was that it was readily available. Innovations had been established in previous years by the evaluator and his colleagues, and the rest of the staff were tolerant of exploration. The evaluator had worked with these people before and was familiar to them. Also, he had experience in the same laboratory course as a designer and demonstrator and had been taught in a similar course as a student by many of the same staff.

Convenience was not the only reason. A challenge was presented by the fact that as previously indicated it would not be possible to use student performance data as one of the measures of course effectiveness. A great deal of co-operation
and effort on the part of the staff involved would have been necessary to make this possible. The experiments were of different styles, and the objectives of the traditional experiments had not been stated. Most importantly the course was so short that the introduction of tests of student performance would have meant removing a large proportion of the course. This was not desired by any of the parties concerned. The evaluation had to take place within this setting without the performance data. This indicated that a novel approach would have to have been adopted and this provided the challenge.

Finally, an intensive investigation had taken place in the previous year by the evaluator. A study had been made of student opinion about the course and information about this had been presented to the course organizers. The staff were interested in the findings, but they were not moved to act on them in any way. In part this can be attributed to the nature of the exercise: it had not been designed to facilitate course changes; and in part to the method of presentation of the results: no expectation of change was announced. The partial failure of this exercise had developed the determination that the course should be evaluated in such a way that information could be made available to staff that was suitable for facilitating changes in the course that could be indicated and implemented easily.

These considerations imposed the initial constraints on the evaluation. There would be no intensive study of student opinion; there would be an emphasis on simple, easily interpretable, measures that demanded little time from staff and students; and the data would be of a kind that could be rapidly processed to enable quick feedback of results to the staff who would have to make the decisions. In addition, one other component was added. A conscious effort would be made to devise measures that were not entirely situation specific and that had potential of being transferred to other courses. This was in anticipation of a future evaluation. (Reported in the next chapter).
Other work in Laboratory Teaching

The literature on laboratory teaching has been extensively reviewed by Tremlett (135). He covered work in the sciences and engineering particularly in Great Britain and the United States. He concludes, "Criticism of laboratory teaching has neither been limited to a single institution, nor even to the educational practice of one particular country. Instead, these criticisms were apparently symptomatic of an approach to tertiary level laboratory teaching internationally".

Much has been written about specific innovations in teaching in the laboratory. See, for example, Read (109), Black Dyson and O'Connor (12), Elton, Hills and O'Connell (49), and Conway and Mendoza (37) on physics laboratory work in the United Kingdom. However, very little has been the subject of evaluation, although both Black, Dyson and O'Connor, and Elton, Hills and O'Connell (49) have included records of student comment.

Rather more has been done concerning the aims of laboratory teaching. Chambers (35) put the question "what use are practical physics classes?" His study of this led him to look at some of the declared aims of physics laboratory teaching in the United Kingdom. In 1963 he published (34) the results of a survey of the aims of physics laboratory work which asked staff in all the physics departments in the country to rate the importance of eleven aims of their laboratory teaching on a ten point scale for first, second and third year courses. See Fig. 4.1 for the first year results. A follow-up study eight years later also looked at the ratings of students who had completed a physics degree and who were asked to rate the importance of the eleven aims for their course.

Lee(76) engaged in a similar exercise for mechanical engineering laboratories. He concludes (75) "Criticisms of practical work in undergraduate mechanical engineering education arise because the aims and objectives of the different procedures are ill-defined and confused, even in the
<table>
<thead>
<tr>
<th>Rank</th>
<th>Mean Rating</th>
<th>Aims of First Year Physics Laboratory Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.3</td>
<td>1. to foster 'critical awareness' (e.g. the extraction of all information from the data; the avoidance of systematic errors)</td>
</tr>
<tr>
<td>1</td>
<td>7.3</td>
<td>2. to stimulate and maintain the students' interest in physics</td>
</tr>
<tr>
<td>3</td>
<td>6.8</td>
<td>3. to familiarise the students with important instruments, devices and techniques (e.g. C.R.O., transistors, vacuum techniques)</td>
</tr>
<tr>
<td>4</td>
<td>6.7</td>
<td>4. to train them in handling data</td>
</tr>
<tr>
<td>5</td>
<td>6.5</td>
<td>5. to train them in writing reports on experiments</td>
</tr>
<tr>
<td>6</td>
<td>5.9</td>
<td>6. to train them in keeping a day-to-day laboratory notebook.</td>
</tr>
<tr>
<td>8</td>
<td>4.6</td>
<td>7. to enable staff and students to meet and talk informally</td>
</tr>
<tr>
<td>7</td>
<td>5.1</td>
<td>8. to illustrate and drive home material taught in lectures</td>
</tr>
<tr>
<td>9</td>
<td>3.1</td>
<td>9. to train in simple aspects of experimental design</td>
</tr>
<tr>
<td>10</td>
<td>2.7</td>
<td>10. to teach some 'theoretical' material not included in the lectures</td>
</tr>
<tr>
<td>11</td>
<td>1.5</td>
<td>11. to impart manipulative skills (e.g. soldering, glasswork)</td>
</tr>
</tbody>
</table>

Results of Chambers (1963) survey of the aims of first year physics laboratory teaching.

**FIGURE 4.1**
minds of academic staff". This concurs with Tremlett (35) who, without reference to Lee and in a different subject area, chemistry, points out, "faculty views not only did not agree on the same laboratory aims for comparable courses in different institutions, but that disagreement existed within the same institution and even between faculty teaching in the same laboratory class. There was also evidence to suggest marked differences of opinion about the relative importance of aims which were held in common". He concludes almost identically to Lee, "although some limited agreement on certain aims has been identified, it is the author's contention that much of laboratory criticism has arisen because the aims of laboratory experience are ill-defined and confused. It is suggested that faculty need to be much clearer about the aims of the laboratory work they provide."

Tremlett also surveys the literature on student opinion of laboratory work. He deduces, "laboratory aims which students recognised were often at variance with those intended by faculty", and suggests, "the reasons for this are not entirely clear, but are thought by the author to arise from students' inadequate awareness of faculty intentions. Either those intentions are not stated sufficiently explicitly for their recognition by the students, or they are omitted altogether".

4.6 Approach Adopted

The evaluation of the present physics laboratory course adopted a perspective based on the declared aims of laboratory teaching. This approach was stimulated by the initial work by Chambers and by Lee and intended to take their work further by applying it to the evaluation of a particular course. It was hoped that an approach in such terms would meet the two aspirations mentioned in the rationale. Namely, that there would be an outcome of practical use and that there would be some transferability of the method.

A start was made by interviewing all the staff involved in the laboratory, seven in all. These were open-ended interviews around the theme of "what should be the aims of
laboratory work for first year students". An attempt was made to extract a list of the aims of laboratory teaching from these interviews.

These lists together with the aims provided by Chambers, and with some of the Lee aims appropriate to physics were assembled to produce a final list of twenty-three. These are given in Figure 4.11. No attempt was made to ask students what aims they thought should be included in such a list. It was felt that it was most appropriate that the aims should be in terms of a staff viewpoint, rather than a student one as this would give a greater chance for any subsequent findings to be acted upon.

These twenty-three aims provided the framework for the evaluation. Staff and students were asked about the course in terms of them. It was expected that areas for improvement could be deduced from the differences in perception of these aims by the two groups.

4.7 Procedure

A questionnaire was prepared which asked staff and students to rate the importance of each of these aims on the different parts of the laboratory course. They were asked to rate them on a five-point scale from 1 (not an aim) to 5 (very important aim) for the two types of experiment - the traditional and the self-service type. Examples of particular experiments were given in case students did not understand the use of the words traditional and self-service. Staff were asked to rate those experiments that they were sufficiently familiar with to feel competent to judge. Both groups were also asked to rate the importance they would give to each aim in their own ideal first year physics laboratory course. This questionnaire was sent by internal post with a covering letter to the staff who had been involved in the course, and to all the physics students who had taken the course. All replies were anonymous so that a follow-up or a reminder was impossible. In addition a similar
1. to instill confidence in physics
2. to teach basic practical skills
3. to familiarise students with important standard apparatus and measurement techniques
4. to illustrate material taught in lectures
5. to teach the principles and attitudes of doing experimental physics
6. to train students in observation
7. to train students in making deductions from measurements and interpretation of experimental data
8. to use experimental data to solve specific problems
9. to train students in writing reports on experiments
10. to train them in keeping a day-to-day laboratory notebook
11. to train in simple aspects of experimental design
12. to provide closer contacts between students and academic staff
13. to stimulate and maintain students' interest in physics
14. to teach some 'theoretical' material not included in lectures
15. to foster 'critical awareness' (eg. extraction of all information from the data; the avoidance of systematic errors)
16. to develop the students' skill in problem solving in the multi-solution situation
17. to stimulate the conditions in research and development laboratories
18. to provide the student with a stimulant to independent thinking
19. to show the use of 'practicals' as a process of discovery
20. to demonstrate the use of an experimental method as an alternative to the analytical method of solving problems
21. to familiarise the student with the need to communicate technical concepts and solutions
22. to provide motivation for the student to acquire specific knowledge
23. to help the student to bridge the gap between theory and practical
questionnaire was sent to all staff in the physics department asking them to rate the aims in terms of their importance in their ideal of a first year physics laboratory course.

There was a good response from the staff. All seven involved in the laboratory and 87% altogether (n = 20) replied. Sixty per cent (n = 24) of the students replied, so there is a possible source of error here.

4.8 Results and Analysis

The assumptions behind the analysis of results were as follows. These govern the form of presentation:

1. If there is a divergence between the ideal aims of a course as expressed by staff and the aims perceived by staff in the actual course, then an improvement can be facilitated by modifying the course so as to bring its aims closer to the ideal aims.

2. If there is a divergence between the ideal aims of a course as expressed by staff and by students, and if staff feel that their set of aims are to be pursued, then it becomes part of the teaching process to modify the ideal aims of the students so as to bring them into closer agreement with those of the staff.

These assumptions are valid only if there is a substantial agreement on the aims within each group (see section 4.10).

From the returned questionnaires the average rating for each group for each aim were calculated. For ease of discussion these will be referred to as follows:

<table>
<thead>
<tr>
<th>Ideal aims</th>
<th>Aims of actual course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Traditional</td>
</tr>
<tr>
<td></td>
<td>Non-traditional</td>
</tr>
<tr>
<td>a) all teaching staff</td>
<td>$R_{IA}$</td>
</tr>
<tr>
<td>b) laboratory teaching staff</td>
<td>*</td>
</tr>
<tr>
<td>c) students</td>
<td>$R_{IS}$</td>
</tr>
</tbody>
</table>

* included in the total number of staff
Where $R_{ij}$ represents the set of results for the subject of the aims, $i$ and the group of respondents, $j$. $i = I$ (aims for ideal), $T$ (aims for traditional), $N$ (aims for non-traditional, self-service).

$j = A$ (all teaching staff), $L$ (laboratory staff), $S$ (students).

The results were analysed in two ways, both in keeping with the principle of simplicity. Firstly, an overall measure of agreement/disagreement between the various groups was taken. As the two end points were fixed, i.e. no importance, and great importance - the scale was taken to approximate to an interval scale, thus enabling a product-moment correlation coefficient to be calculated between pairs of ratings ($72, 73$). These coefficients are tabled in Fig. 4.2.

Secondly, each aim was considered separately and differences between the ratings given by various groups were studied. A form of presentation was chosen which, although it eroded the original data a little, produced a display that facilitated easy discussion of the differences between groups. For each group the means for each aim were converted to a rank order. The ranking of aims was then plotted graphically to show the differences between the groups of staff and students for each component of the course, and also the rating of ideal aims. Staff and student's ranking of (a) the ideal aims of a first year laboratory course, (b) the aims of the traditional experiments and (c) the aims of the non-traditional experiments are displayed in Figs. 4.3, 4.4, 4.5.
FIG. 4.2

<table>
<thead>
<tr>
<th></th>
<th>$R_{TL}$</th>
<th>$R_{NL}$</th>
<th>$R_{TS}$</th>
<th>$R_{TS}$</th>
<th>$R_{NS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{IA}$</td>
<td>0.61</td>
<td>0.39</td>
<td>0.66</td>
<td>0.69</td>
<td>0.38</td>
</tr>
<tr>
<td>$R_{TL}$</td>
<td>0.17</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{NL}$</td>
<td></td>
<td></td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{TS}$</td>
<td></td>
<td></td>
<td></td>
<td>0.33</td>
<td>0.65</td>
</tr>
<tr>
<td>$R_{TS}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.31</td>
</tr>
</tbody>
</table>

Product moment correlation coefficients between means for each rating
Staff and student ranking of ideal aims.

Fig. 4.3
Staff and student ranking of the aims of traditional experiments.

*Fig. 4.4*
Fig. 4.5

Staff and student ranking of the aims of non-traditional experiments.
Perfect agreement between staff and students would be represented by each aim lying on a line passing through the origin at 45°. This is shown and labelled the line of agreement. A measure of the disparity between the two groups can be obtained by examining the distance from the line of each aim. For example, in Figure 4.3 aims 9, 10 and 13 all lie at a distance equivalent to a difference of greater than 10 places on the ranking scale. In Figure 4.4 aims 9, 17, 19, 21 and 23 all fall outside this limit, and in Figure 4.5 only aim 21 is outside. This tends to indicate that there is most agreement between staff and students on the aims of the non-traditional experiments and least agreement for traditional experiments, and using this measure of extremes, the ideal aims fall between the two. This agrees with the results of the general measures provided by the correlation coefficients where the respective coefficients for non-traditional, traditional and ideal are 0.67, 0.23 and 0.66.

Figures 4.6, 4.7, 4.8 and 4.9 give an alternative presentation of the rankings of aims and it is from these graphs that additional information about discrepancies between aims is derived. Agreement between staff and students, and between ideal aims and aims of the actual courses would in each case be represented by points at the origin of axes. In practice statistical errors would require this point to be replaced by a circular area centred at the origin. Individual aims can be examined, therefore, in terms of their distance from the origin.

In Figures 4.6 and 4.7 aims situated in the first quadrant have been rated as being of less importance than the ideal for both staff and students i.e. both groups rate these aims as being under-represented in the present course. For aims in the third quadrant the opposite would be true - both staff and students agree that a particular aim has greater emphasis in the present course than the ideal.
Differences in the aims of traditional experiments for staff and students relative to each group's conception of ideal aims.

Fig. 4.6
Differences in the ranking of the aims of non-traditional experiments for staff and students relative to each group's conception of ideal aims.

Fig. 4.7
Differences in the ranking of the ideal aims and the aims of the traditional experiments, for staff and students.

Fig. 4.8
Fig. 4.9

Differences in the ranking of the ideal aims and the aims of the non-traditional experiments, for staff and students.
In the second and fourth quadrants aims are found on which there is a disagreement between staff and students on whether the aim is under-represented or over-represented in the present course. In the fourth quadrant for example, there would be aims which students would agree to be over-represented and staff under-represented in the course.

In Figures 4.8 and 4.9 aims situated on the horizontal axis are those on which staff and students agree in the ideal case, and aims on the vertical axis are those on which staff and students agree for either the traditional, (a), or non-traditional, (b), experiments. Therefore, aims in the first quadrant are those which staff thought were more important than students for both the ideal and present courses, and vice versa for the third quadrant. The fourth quadrant would, for example, include aims which students would agree to be more important than staff in the ideal cases, but which they thought were less important than staff in the actual course they received.

This form of graphical representation allows for discussion of individual aims, and can be used to deduce detailed information about the course from the points of view of staff and students.

Examination of a single aim will provide an example of the procedure that may be adopted. Take aim 1, "to instill confidence in physics", it is rated of equal importance by staff and students in an ideal course, 11th or 12th out of 23 (Figure 4.3). In the traditional experiments it is rated more highly by staff i.e. 12th, than students, 19th (Figure 4.4), and in the non-traditional experiments the difference is even greater - 5th as opposed to 15th (Figure 4.5).

Turning to Figure 4.6, aim 1 is situated on the student ordinate, i.e. staff give it the same ranking for both the traditional component and the ideal, students, however, regard it as being under-represented in the traditional course
relative to their ideal by 8 places. For self-service experiments, Figure 4.7, staff think aim 1 to be over-represented in these experiments relative to their ideal, 7 places, but students again think it to be under-represented, though only by 4 places.

In Figure 4.8 aim 1 lies close to the horizontal axis indicating that staff and students agree on its relative importance in an ideal course, but in the traditional part of the course students think it under-represented compared to staff by 7 places. For the self-service experiments, Figure 4.9, there is a similar finding – students think aim 1 is under-represented in this component by 10 places.

The information provided can be used in many ways for the purposes of course improvement. If there is high agreement on the aims within each group, then a discrepancy between the ratings of any given aim by the two groups, or between the ratings of the aims of a part of a course and the ideal aims, suggest a need to study the particular aim more closely. This may require the transmission of the significance of an aim from staff to students or vice versa, it may involve a modification of a part of a course so that an aim is more apparent and it may require discussion amongst staff, and between staff and students to clarify what the discrepancy is thought to be due to and what changes can be made on the basis of this. In this study there is high agreement between staff and students on the ideal aims of a course, so that potentially improvements can be made from the points of view of both groups.

4.9 Outcomes

The preceding section outlined a method for handling the results of an aims questionnaire to provide information for course improvement. These ideas and suggestions were incorporated in a report to the staff concerned with running the course. The report produced reactions from many of them.
The reactions were in terms of ambiguities in the wording of the report and minor inaccuracies on points of detail. The author had discussions with the staff involved in the course. During these little or no mention was made of the use to which the results could be put. In short this study resulted in no action to improve the course that was being investigated.

However, it did meet its original aim of providing an evaluation of the two types of experiment from the same perspective. A method was developed that assessed the staff and students perceptions of both styles of experiment in terms of the established aims of physics laboratory work.

The outcomes of the study were thus both positive and negative. Evidence was obtained about the course, but when presented to the staff it did not lead directly to course improvements.

Factors which influence the implementation of evaluation findings are crucial to any study. It is a waste of time and resources to engage in an activity designed to produce change if no change results. It may be though, that the study of a failure of this kind can illuminate some basic problems in the evaluation process.

An analysis of the factors behind this outcome cannot be objectively made by the present author. However, some factors seemed pertinent at the time, and three years later many of them still seem important.

The two important factors that emerge from this study were:

1. The evaluation was not sponsored directly, or initiated by those people who were responsible for implementing possible results. It was initiated and carried out by an external agency i.e. from an evaluator from another department in the university. The staff involved in implementation were ready guinea pigs for the research
that was done on them, but they were not in the position of control or active influence on the strategies and procedures that were adopted.

2. The consumers of the research, namely, the staff involved in implementation, were not committed to any change on the basis of the evaluation findings. In this particular course the innovative staff were not in the position of decision makers. They were in the position of being allowed to do only what the controlling staff consented to. It may be, if the controlling staff were presented with a new experiment that could easily be substituted for an existing one, they would allow the change. They were not committed to be personally involved in the change process, however.

The second point is confirmed by Dow (152) who has stressed the necessity of prior commitment to change on the part of the agency who were the consumers of the evaluation. In both the case of Dow and the present study the consumers were university teachers.

These factors emerged from the failings of an empirical investigation, but they can also be seen from an ethical standpoint. They raise the questions 'Should formative evaluation be tolerated in a situation where it is imposed on the prospective consumers of it'? 'Should evaluation be used as a covert strategy to impose change on those unwilling to engage in change activities'? These issues did not arise in the form that they are stated here because the evaluation agency did not have sufficient power or control over the situation. They may well arise if course evaluation is prescribed by authority.

Despite the apparent failure of the study to meet its own main aim, that of effecting course improvement on the course that was being evaluated, there were some subsidiary outcomes. The first was empirical, and the second relates
to a theoretical issue concerning models of evaluation.

Firstly, the evaluation results reported here were used to effect a course improvement. A year later the self-service experiments were removed from the laboratory course without explanation. They were immediately adopted by another department who were involved in the same common course. In effect this meant that the same students did self-service experiments, but in a different course. This change allowed the innovative staff to restructure the part of the course that was allocated to these experiments. They took as their basis the results of the evaluation that related to the self-service, or non-traditional experiments. Aims for the sub-course were selected from those aims that both staff and students rated highly for the self-service experiments, and for the ideal course. Specific objectives for individual experiments were related to these overall aims and students were invited to design their own course consisting of self-service experiments by selecting experiments which gave a good coverage of the overall course aims. An evaluation of this system is reported by O'Connell, Penton and Boud (95).

Secondly, it was found that there was a high degree of agreement both between staff and students, and between staff. This is in apparent contradiction to the statements of Lee (75) and Tremlett (135) quoted in section 4.5. However, it is important to look more closely at the contradiction. If we do so we find that the term "aims" is not being used in the same closely defined fashion. In the chapter on models of evaluation, hierarchies of aims and objectives were discussed (70), and it was pointed out that it is more likely that agreement about general aims would be present, than for specific objectives. This implies that at one level of discussion a group of, say, university teachers would agree about the aims of a course, but at another level they may well disagree. If this is the case then it is important for evaluation activities to be directed at the level of aims and objectives where there is agreement between members of a given
group, but dissention between groups. Under these conditions fruitful discussion between groups is possible, and attention is drawn to real differences rather than differences in emphasis.

4.10 Statistical Footnote

One of the assumptions of the analysis in section 4 was that there was a high level of agreement between the rating of each group. This is an assumption that needs to be tested. It also leads to a consideration of two properties of a questionnaire that should be examined before reliance is placed on the results that are produced. These are reliability and validity:

4.10.1 Reliability

The reliability of a scale is a measure of the consistency with which the scale measures a property. Various measures of reliability are available (52); in this case a form called split-half reliability was used. This enables an estimate of reliability to be obtained from a measurement at one point in time.

The results for each group on the aims questionnaire were randomly divided into two, and a product-moment correlation coefficient, \( r \), was calculated for the ratings on each pair of samples. From this a split-half reliability coefficient, \( R \), was derived using the Spearman-Brown formula (52):

\[
R = \frac{2r}{1+r}
\]

The results of these calculations for both groups are shown in Figure 4.10.

In addition to this an additional check on the reliability was obtained for the staff. Each member of staff was asked on the questionnaire to state whether they regarded themselves
Figure 4.10

<table>
<thead>
<tr>
<th></th>
<th>SAMPLE A</th>
<th></th>
<th>SAMPLE B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>$r$</td>
<td>$R$</td>
<td>$r$</td>
<td>$R$</td>
</tr>
<tr>
<td>$R_{\text{IA}}$</td>
<td>0.74</td>
<td>0.85</td>
<td>0.67</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Staff $(n=20)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{\text{IS}}$</td>
<td>0.66</td>
<td>0.79</td>
<td>0.60</td>
<td>0.75</td>
</tr>
<tr>
<td>$R_{\text{TS}}$</td>
<td>0.53</td>
<td>0.69</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>$R_{\text{NS}}$</td>
<td>0.60</td>
<td>0.75</td>
<td>0.53</td>
<td>0.69</td>
</tr>
<tr>
<td>Students $(n = 24)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Product moment correlation coefficients and reliability coefficients for random split-half groupings.
as primarily experimentalists, or theoreticians. It would be expected a-priori that there would be some difference in the ratings of the aims of an experimental course between these two groups. The means for each group were calculated as before, and the product moment correlation coefficient of 0.88 was obtained between these means. This gives a reliability coefficient $R = 0.94$.

4.10.2 Validity

A common way of assessing the validity of a set of educational aims or objectives is to get a group of outside judges to examine the list of aims which have been prepared and to ask them to assess this list in terms of whether they are meaningful and consistent aims for the type of course under discussion (114). In the present case, however, this method is not applicable. Some of the aims were taken from previous interviews with the same staff who rated them in the questionnaire, whilst others were collected from the work of Chambers and Lee both of whom had derived them from similar sources. The aims in this context are in a sense self validated so that external judgement would therefore be inappropriate.

A partial check on the completeness of the list of 23 aims for the particular groups that rated them was made though. On each questionnaire in addition to the 23 aims already mentioned, two spaces were left at the end of the list and respondents were given the opportunity to fill these gaps with aims of their own choice. In the staff group only two gave additional aims, and in the student group only three included other aims, of which only two were considered to be at all serious. Two physicists independently of the author examined these aims. They thought them to be not significantly different from aims already on the list.
4.10.5 Other evidence of Reliability

The laboratory aims questionnaires has been tried in other circumstances to the one described. In two of these situations it was possible to collect evidence pertinent to the reliability of the scales. These investigations took place in (a) the first year physics laboratory course of a new university, and (b) the first year physics laboratory course for students taking physics as a minor subject in a provincial university. These will be referred to as course A and course B.

4.10.(3)1 Course A

An aims questionnaire was administered to all first year and second year students in the department at the end of the spring term. First year students at this time had completed two terms of a physics laboratory course, second year students had completed the same period in the second year laboratory. It had been two terms since the second year students had experienced the first year course which was to be examined. Amongst other things, both groups were asked to rate the importance of the same list of aims of laboratory teaching on the same scales as before. Namely, the importance of aims in their actual first year laboratory course, and the importance these aims would adopt in their ideal first year laboratory course. A report of the findings of this investigation is available elsewhere in an internal report. However, two aspects relate to the reliability of the aims questionnaire scale. Firstly, what do the split-half coefficients show, and secondly, is there any difference between the first year students ratings and those of the second year students looking back on essentially the same course?

Figure 4.11 shows the correlation and reliability results for two split-half random samples. Figure 4.12 shows the correlations between first and second year ratings for both the actual and ideal courses.
FIGURE 4.11

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>R</th>
<th>r</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>0.93</td>
<td>0.79</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>0.87</td>
<td>0.93</td>
<td>0.93</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>0.91</td>
<td>0.91</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>0.94</td>
<td>0.97</td>
<td>0.93</td>
<td>0.96</td>
<td></td>
</tr>
</tbody>
</table>

Course A: Random split-half correlations and reliability coefficients

FIGURE 4.12

<table>
<thead>
<tr>
<th>Rating</th>
<th>Ideal Course</th>
<th>Present Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>R</td>
</tr>
<tr>
<td>0.92</td>
<td>0.96</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Course A: Correlation coefficients and reliability coefficients between mean ratings of first year students and second year students rating the first year laboratory course.

FIGURE 4.13

<table>
<thead>
<tr>
<th>Present Course</th>
<th>Ideal Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
</tr>
<tr>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>0.86</td>
<td>0.92</td>
</tr>
<tr>
<td>0.92</td>
<td>0.96-</td>
</tr>
</tbody>
</table>

Course B: Random split-half correlations and reliability coefficients.
These figures support those found for the original aims questionnaire results. It may reasonably be assumed that the questionnaire is a reliable measure in Course A.

4.10.(3)2 Course B

In this case an aims questionnaire was administered two years running to the students in a first year subsidiary physics laboratory course. During this time changes had been made to the subject content of the course in order to up-date it. However, no change had been attempted in the aims for the course. Students were asked to rate the actual laboratory course they were attending, and also their ideal first year physics laboratory course in terms of the importance of the aims in the aims questionnaire.

For both years split-half reliability coefficients were calculated. In 1972, though, instead of dividing the groups randomly, they were chosen on the basis of their major subject: chemistry, mathematics and mining engineering. Coefficients ranged from 0.88 to 0.98 for both the actual course, and students rating of the ideal. In addition, a comparison between the ratings in 1972 and 1973 was made. Correlation coefficients and reliability coefficients were calculated between these two ratings. They were found to be high. See Figure 4.15.

The evidence from both Course A and Course B supports the assumption of the reliability of the aims questionnaire scales. In both courses the questionnaire was administered in a normal laboratory period, and the response rates were 100% of those students attending the laboratory on that day. This represents 90-95% of the total possible population. No evidence was obtainable from faculty as the numbers were too small to validly use correlation measures.
FIGURE 4.14

<table>
<thead>
<tr>
<th>Actual Course</th>
<th>Ideal Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>R</td>
</tr>
<tr>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>0.90</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Correlations and reliability coefficients for comparison between major subject groupings Course B.

FIGURE 4.15

<table>
<thead>
<tr>
<th>Actual Course</th>
<th>Ideal Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>r</td>
</tr>
<tr>
<td>0.87</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Course B: Comparison between 1972 and 1973 student ratings.
Although the evaluation described in this chapter failed to meet its original stated aim; to provide information that would lead to course improvement, there have been several important outcomes. Firstly, a technique has been developed which can reliably provide information about how different groups of people perceive the aims of laboratory teaching, and procedures for presenting and handling this information have been tried with some success. Secondly, some important issues concerning the implementation of evaluation studies have been raised.

However, there is still substantial room for improvement in measures of this type. Some have been mentioned in connection with the sponsors commitment to evaluation and these relate to any evaluation strategy. More specifically, improvements can be made in the details of the method. These include:

1. Student aims should be included in a survey of aims of laboratory work. In the present study all aims have been derived from a faculty viewpoint, they are stated in the terms that they would state them. This is, perhaps, valid in a situation where decision making lies firmly with faculty members. If a wider view is taken of teaching/learning activities, students would be involved in the discussion and decision making processes in course changes. Under these circumstances, an attempt should be made to assemble information from both standpoints, without initial bias in the research instruments.

2. Additional information needs to be gathered which will assist in the interpretation of aims questionnaire information. This has been done in the studies related to Course A and Course B. Aims questionnaire results pinpoint areas of difficulty, but they do not give a prognosis for the situation. Clues to this can be provided by ancillary open-ended questions about the
course. The data is not so easy to handle, but its richness can more than compensate for this. Diversity of research tools is a good remedy for flaws and limitations of individual methods.

Finally, it might be considered, in this evaluation study, whether some completely different approach would have been more suitable. Answers to this question are mere speculation; any measurement affects the system that is being measured; it cannot be said what might otherwise have been.
5. CASE STUDY: INTRODUCTORY BIOLOGY LECTURE COURSE

5.1 Introduction

Unlike the previous chapter, this chapter presents an evaluation study of a course that was without any innovative component. It is an introductory lecture course in cell biology and genetics. The course is given entirely through lectures, no tutorial contact or other ancillary contact, for example, in laboratories, is provided. It was these very features that encouraged the present evaluation activities to be centred on this course.

The aim of the investigation reported in this chapter is: to find measures that can generate useful evaluative information to be derived from, and take place within the constraints of, a traditional lecture course and to enable course improvements to be made. This information should satisfy the criteria of usefulness for planning course changes.

5.2 The course

The course investigated was a two term lecture course of 35 hours duration in the first two terms of a common course for first year students. Students taking the course were following degree courses in Human Biology, Microbiology, Biochemistry and Nutrition. They followed a common introductory course in subjects that formed the foundations of these subjects; Cell Biology and Genetics was one of them. It was the major introductory biology course for all these students. The number enrolled in the course was around one hundred.

The student group had a diverse background as far as pre-knowledge of the subject was concerned. Most students had at least one biological subject at 'A' level, many had two biological subjects, but around 10% had done no biology beyond 'O' level at all.
The course consisted entirely of lectures, there was no formal additional staff-student contact for this subject. Students were invited by the lecturer to discuss their academic problems with him personally if they wanted to. Very few (<5%) took advantage of this opportunity. However, at the end of each lecture a few (rarely more than four) students stayed behind to ask questions of the lecturer. Thus, for the majority of students the course consisted of lectures, plus the exam held at the end of the vacation after the course.

The nature of the objectives that the lecturer had for the course will be mentioned in detail later. They were, in general, of an attitudinal type. That is, they aimed not at the acquisition of specific knowledge, but rather at an appreciation of the methodology of biology, and at developing in the students a thoughtful and inquisitive view of biology. The lecturer considered that it was not essential for students to learn any particular given facts, and he expected different students to acquire a detailed knowledge of different parts of the course, and he hoped that they would all develop the attitudes outlined above.

There were no novel features in the course. An overhead projector was used instead of a blackboard, but this was the greatest departure from what can be called a traditional lecturing style: the lecturer addressing from a dias a tiered lecture theatre filled with around 100 students.

Neither were there any overt problems with the course. The lecturer had given the same or a similar course for many years and was not dissatisfied with it in general. Minor modifications were made each year to up-date subject matter, but no other changes were anticipated or suspected. Student reaction also appeared to be favourable. No criticisms had been expressed through any channels, formal or informal, known to the lecturer. In summary it can be said that the course appeared to satisfy all concerned.
Specific Constraints

The fact that the course was run as a traditional lecture course places a great constraint on the evaluation. Firstly, there is only a limited period of contact with the students, and this contact is in a prescribed form, i.e. a lecture. So any form of evaluation that would involve, for example, a regular check on student progress, or feedback from students during the course, would change the form of the course. A constraint was accepted that if a lecture course is being examined an evaluation strategy must be chosen that is compatible with that method and does not change it into something else. At the same time it is desired that the maximum possible information should be obtained within the limited time available.

Secondly, there is the constraint of the examination. A traditional essay-type examination was given three to four weeks after the end of the course. Students were required to answer three questions out of six in two hours. If the assumption is made that the six questions represent a full coverage of the course (which they manifestly do not), then a student can answer three questions and get full marks. This represents 50% coverage of the course. If a pass mark of 40% is assumed as typical (which it was), then a passing grade can be obtained by a student who demonstrably knows a maximum of 20% of the course material. As the initial assumption about the examination representing a full coverage of the course can very easily be questioned, then it suggests that even under the chartible conditions outlined the examination mark can be a poor measure of the students' attainment. If, for this course, one takes into account the attitudinal rather than syllabus orientated objectives and examines the examination papers and compares them to the established aims of the course then more doubt is cast on the examination as a valid measure of student performance. The papers in the biology course tend to test relatively low-level cognitive objectives, a common finding in science examinations (see for example (11)). So we have the
constraint on the evaluation that the only build-in measurement is an inadequate measure of the kind of attitudinal objectives that the lecturer is trying to achieve. The immediate suggestion is that the examination should be changed. This was not possible in the situation as many other decision-making committees would have to have approved in time.

Other constraints revolve around the relationship between the lecturer and the evaluator, and the expectations of the lecturer on his agreeing to co-operate in an evaluation study of his course. Firstly, the evaluator was not called in by the lecturer to facilitate course improvements. The evaluator approached the lecturer and asked to study his course for research purposes. The aim of the research was explained, and was slightly modified after discussions. The lecturer was quite willing for the study to take place so long as it did not disrupt the course or take up too much time. This leads to the second constraint. The lecturer was interested in the study, but was not committed to make any changes. He had not asked to be evaluated and had no obligation to act on any findings.

These constraints were taken into account in the research strategy that was adopted.

5.4 Rationale for Choice

It has been explained that this study was sought by the evaluator and was not commissioned directly by a teacher. Why, it can be asked, was this particular course chosen, other than for the obvious reason of the presence of a co-operative teacher?

The prime consideration was that the course was not innovative and was therefore more typical of university courses than some experimental ones that were available. It did not have any overt problems, so it was anticipated that it was probable that any areas or change that were pinpointed as a result of the study would not be major ones. It was also a
course whose lecturer aspired to attitudinal objectives. This introduced problems in the evaluation strategy that would require novel solutions to overcome them. The testing of cognitive objectives is well established (15), the measurement of the achievement of affective objectives in university courses is a much more unknown area.

Taking into account the constraints outlined in the previous section and the above considerations a research strategy was adopted: based on the results of a pilot investigation started the previous year. This had aimed to look at the feasibility of obtaining student feedback on various aspects of the course.

5.5 Pilot background Study

The pilot study covered many areas that were not covered in the evaluation that is reported here. A full description of it would divert attention from the development of the strategy that was finally adopted. Certain specific outcomes will be discussed here, but it is relevant to note that the choice of these has been post-hoc. Some lines were investigated that were dropped at an early stage. The major reason for this was that a continuation and development of them would fall outside the self-imposed constraints of the course.

5.5.1 Student Aims

One of the problems that was studied in the pilot run was the extent to which students perceived the lecturer's aims for the course. Mid-way through the first term of the course students were posted a questionnaire. One of the questions on this was "What do you think are the main aims of (the lecturer) in giving the course?" Students were encouraged to give at least three aims. These aims were content analysed into eleven main categories. Just before the end of the course students were presented with another questionnaire. One of the questions asked them to rate each of these eleven
aims in terms of the degree to which they agreed that these were aims of the lecturer. A five point scale was provided from 1 (strongly disagree) to 5 (strongly agree). The results are shown in Fig. 5.0.

These results were discussed with the course lecturer. He expressed interest in them, but said that they were not useful to him in terms of improving the course. Some of the aims were ones he would have chosen himself, for example, "to show that some concepts in biology are not definite and proven"; others were ones he would tend to reject, for example, "to teach for the exam".

For this reason the approach to evaluation through obtaining student perceptions of student aims was not continued. It did not provide information that the lecturer could use, and in some cases the resultant aims were somewhat vague and difficult to interpret. An example of an aim of this type is, "to describe a way of looking at biology".

5.6 Approach Adopted

In this course an approach was adopted which took a lecturer-centred stance. It was considered that if the lecturer was the consumer of the evaluation, then a perspective based upon his framework of the course would be most appropriate.

Also, if the course were to be improved from this standpoint then this should be done relative to the lecturer's criteria for the success of this course. This is similar to criterion-referenced testing which has been developed to measure student performance in terms of pre-specified criteria, (54) as distinct from norm-referenced measures that specify outcomes in terms of the differences between students, or as in this case the difference between the given course and other courses currently being studied by the same students. Such an approach is non-comparative, it measures against given criteria.
<table>
<thead>
<tr>
<th>Aim</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>to bring everyone up to the same standard</td>
<td>3.5</td>
<td>1.1</td>
</tr>
<tr>
<td>to teach certain specific topics in biology</td>
<td>3.7</td>
<td>1.0</td>
</tr>
<tr>
<td>to describe a way of looking at biology</td>
<td>3.4</td>
<td>1.2</td>
</tr>
<tr>
<td>to show that some concepts in biology are not definite and proven</td>
<td>4.0</td>
<td>0.9</td>
</tr>
<tr>
<td>to stimulate, and create interest in the subject</td>
<td>3.6</td>
<td>1.2</td>
</tr>
<tr>
<td>to teach for the exam</td>
<td>2.8</td>
<td>1.3</td>
</tr>
<tr>
<td>to show the relevance of the subject to one's main degree studies</td>
<td>2.6</td>
<td>1.1</td>
</tr>
<tr>
<td>to familiarise us with the ways in which biologists think</td>
<td>3.2</td>
<td>1.2</td>
</tr>
<tr>
<td>to encourage us to look critically at the subject</td>
<td>3.6</td>
<td>1.1</td>
</tr>
<tr>
<td>to relate biology to basic physics and chemistry</td>
<td>1.9</td>
<td>1.0</td>
</tr>
<tr>
<td>to encourage us to read books and articles</td>
<td>3.4</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Student ratings of what they perceive to be the main aims of the course lecturer.

*Fig. 5.0*
The strategy that was finally adopted had two components and relied on the pre-specification of the lecturer's aims for the course, and on his hoped-for affective outcomes. These were developed into an aims questionnaire, modified from the version discussed in Chapter Four; and an attitude questionnaire based upon the idea of the semantic differential (96).

5.6.1 Aims Questionnaire

The first stage was the determination of the aims that the lecturer had for the course. Lengthy discussions with the lecturer over a period of several months produced a final list of 15 aims. An attempt was made to specify more detailed objectives for some parts of the course. This attempt was abandoned at an early stage as the time involved became too great and the number of objectives too lengthy. For example, an attempt to analyse the objectives of a single one hour lecture produced over one hundred specific objectives and took many hours to determine.

Further attempt at this kind of analysis were abandoned as being too time consuming relative to the use to which the results could be put.

The 15 aims were considered by the lecturer to fully represent what he was attempting to do in the course. They were written in terms of lecturer activity rather than as objectives describing desired student performance. This was a means of constructing an evaluation from the lecturer's frame of reference. The lecturer did not think of his teaching directly in terms of what students would be able to do as a result of the course.

These aims were assembled into a questionnaire. Students were asked to rate the aims on four separate scales to produce a profile of students' perception of the course in terms of the lecturer's framework. They were instructed that these aims
were possible aims for a course of the type they were following. They were not told that they had been constructed by the course lecturer.

The four scales were as follows:

A. the importance of aims for the present course — on a scale from 1 (not important) to 5 (very important).

B. the importance of aims in a course which the student would consider ideal — on the same scale.

C. the probability of the aims being examined — on the scale from 1 (not likely) to 5 (very likely).

D. the degree of success of the course in achieving the aims — on the scale from 1 (not successful) to 5 (very successful).

The questionnaire was administered in the penultimate week of the course. Questionnaires were returned through the internal post and a final response rate of 78% was obtained.

The same questionnaire, with some additions, was similarly administered in the following year to the group of students then engaged in the course. It was completed in a lecture period, when a lecture could not be given, and returned immediately. It took approximately ten minutes to complete. The response rate was 98% of those attending the lecture — 86% of the students enrolled on the course.

5.6.2 Pre and Post-attitude measure

A style of questionnaire similar to that used for the semantic differential (96) was adopted. In this students are asked to rate various concepts on a 7-point scale, the poles of which were bi-polar adjectives. For each concept many scales are provided. In this way it is intended to get a comprehensive profile of the students' view of each concept.
The measures used here differed from the original semantic differential in one substantial way. Each scale was intended as an evaluative scale. And adjectives prescribing each scale was chosen by the course lecturer as the important dimensions of possible affective change in the course. Some of these scales were coincidentally the same as those used by Osgood (96), others were taken from the work of the PEEP project (101), and yet others were generated by the lecturer. The lecturer indicated the hoped for direction of change in each scale.

A questionnaire comprising these concepts and scales was administered at the start of the course and again in the penultimate week of this course. It was intended that these should represent a measure of students pre and post-course attitude towards the course and certain concepts covered in the course. This was the same as the use to which the PEEP project (102), and the IPI project (80) put the semantic differential. No reports as to the usefulness of these measures to these other projects are available at the time of writing, however.

In the first year of the study the concepts included were: My expectations towards the course/looking back on the course; Genetics; Physico-chemical explanations of living processes; the Gene; Cytoplasm; and D.N.A. In the second year the questionnaire was shortened to reduce the time taken for it to be filled in. The total number of scales was reduced and the concepts: physico-chemical explanations of living processes and Genetics were removed. The second version of the questionnaire can be found in Appendix I.

5.6.3 Ancillary Measures

- In the first year of the study the aims questionnaire and the pre and post-attitude tests were used alone. In the second year it was considered necessary to get some subject content specific information. That is, what areas of cognitive content of the course were being little understood by the students?
The question was approached by constructing a scale that included a list of the main subjects covered in the course, in the order that they were covered: structure of protein molecules, electron microscopy of cells, etc. Students were asked to rate their degree of understanding of each of these topics on a five-point scale. The end points of the scales were operationalized in terms of whether or not students would choose to answer an examination question on the topic. The scale ranged from 1 (do not understand - would not attempt question) to 5 (understand very well - would choose question on it).

Also included with this scale was an open-ended question designed to check on the validity of the scales and to allow students to express themselves fully concerning the course without the constraint of a fixed number of possible responses. The question asked: "Please comment on any aspects of the course you wish, thinking particularly of those parts/methods that you feel could be improved".

Finally a check was made on students' reaction to filling in the complete questionnaire, which now included the aims questionnaire, the scale of understanding of course content and the open-ended question on course improvements (the attitude test was given separately). They were asked: "Please give your reactions to filling in this questionnaire". The complete questionnaire is included in Appendix II.

5.7 Results and Analysis

5.7.1 The Aims Questionnaire

The rationale for the type of analysis that the results were subjected to was as follows:

It is assumed, that in the most satisfactory course, students should feel that the various manifestations of the aims of the course should be, and should be seen to be,
consistent. That is, what they think is being achieved should be congruent with what they hope the course should achieve, with what they think the course will achieve, and what they think will be assessed at the end of the course.

It is also assumed that this can be estimated by measuring the extent to which students see agreement between the aims for their ideal course, the aims of the present course, the degree of achievement of the aims, and the likelihood of the aims being examined.

Four comparisons between these scales were made. They were:

1. the importance of the given aims in the present course - the importance of these aims in the students' ideal course (A - B).

2. the importance of the aims in the present course - the probability of these aims being examined (A - C).

3. the importance of the aims in the present course - the degree of achievement of them in the present course (A - D).

4. the probability of these aims being examined - the degree of achievement of them in the course (C - D).

These comparisons are presented in Figures 5.1 and 5.2 to display the differences in ratings directly. The ends of the bars represent the mean rating of the students' score on the five point scale. The lengths of the bars represent the differences between the mean ratings for each comparison. These displays are in keeping with the principle outlined in Chapter Four of simplicity of presentation in order to highlight problem areas. From them it is possible to see directly those cases where there is very low congruence between the ratings of an aim on different scales.
For each pair of ratings a Student - t statistic was computed to show the statistical significance of the difference between means. This was done to give a guide to degree of separation of the two sets of ratings as it takes into account the standard deviation as well as the mean. An assumption was made that the scale approximated to an interval scale so that calculation of the t-statistic is valid (72). Differences significant at 5%, 1% and 0.1% levels of probability are displayed in Figures 5.1 and 5.2.

The results can be used diagnostically. Each aim can be examined in turn and the nature and extent of the incongruences between the ratings on different scales can be examined. These can provide clues for the way students see the problems of the course. They are systematic clues and cover those aspects that the lecturer considers of importance. Unlike the laboratory aims questionnaire, the method of construction of the aims has ensured that they are all high priority aims so no weighting in the analysis need be given to make any one aim of greater importance than any other.

For example, taking aim 15 in the 1971-72 course (Fig.5.1), "to give a student an interest to investigate the subject further". Firstly, it is given the highest rating of any aim in the students' ideal course. Secondly, it is rated slightly less importantly for the present course. It has the fourth greatest difference of any aim between the present and ideal ratings. Thirdly, its rating in terms of the success of the course in achieving the aim is much lower. It is the sixth lowest aim on this scale. Finally, the probability of this aim being examined is rated substantially lower still. It is the second lowest aim on this scale, and the difference between its rating on this scale and on the scale representing its perceived importance in the present course is huge: it is the largest discrepancy of all.

It can be concluded that in terms of the aims questionnaire scales aim 15 is seen as important by the students both ideally,
THE DIFFERENCE IN THE RATING OF THE AIMS OF THE COURSE IN TERMS OF (A) THEIR IMPORTANCE IN THE
PRESENT COURSE, (B) THEIR IMPORTANCE IN THE STUDENTS' IDEAL COURSE, (C) THE PROBABILITY OF BEING
EXAMINED, AND (D) THE DEGREE OF SUCCESS IN THE PRESENT COURSE.

<table>
<thead>
<tr>
<th>AIM</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. to expose a student to a minimal body of factual information which is essential to his future work in this or other parts of the course</td>
<td>★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. to integrate a recapitulation of 'A' level into the new material</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. to stimulate a student to read recommended literature</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. to inform a student of sufficient concepts and terminology to enable him to gain further knowledge from the recommended literature</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. to stimulate reading outside the lecture content</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. to encourage a student to accept that statements about physics and chemistry may be taken as a starting point for the purposes of constructing biological hypotheses</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. to give examples of hypotheses which may be advanced to relate biological phenomena to accepted 'facts' about the properties and behaviour of molecules</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. to give examples of observations and experiments which demonstrate how certain experiments constitute evidence for a particular experiment</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. to show students that biology is a developing science in which new observations and concepts are being made and older ideas modified</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. to enable a student to make criticisms and discern gaps in knowledge or explanations in what he reads</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. to enable a student to suggest (albeit wrongly) alternative ideas to those found in literature he has read</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. to make a student aware of the diversity and variability of biological phenomena, and of the different types of explanations that may be put forward to interrelate them</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. to lead a student to be aware that descriptions of all structure and function are conceptual models or hypotheses, that have been suggested by observational or experimental results, so that statements of 'fact' need to be qualified accordingly</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. to stimulate the student to attempt to understand new discoveries and interpretations as they occur</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. to give a student an interest in exploring the subject further</td>
<td>★★★</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MEANS ON A 5-POINT SCALE

For each aim the four blocks represent the difference between ratings on each scale, shown below:

- A-B: the importance of aims in the present course
- A-C: the importance of aims in the students' ideal course
- A-D: the probability of aims being examined
- C-D: the degree of success in the present course

The arrows represent the direction of positive difference between the mean ratings on each scale.

P < 5%
PP < 1%
PPP < 0.1%

Fig. 5.1
THE DIFFERENCE IN THE RATING OF THE AIMS OF THE COURSE IN TERMS OF:
1. THEIR IMPORTANCE IN THE PRESENT COURSE,
2. THEIR IMPORTANCE IN THE STUDENTS' IDEAL COURSE,
3. THE PROBABILITY OF BEING EXAMINED,
4. THE DEGREE OF SUCCESS IN THE PRESENT COURSE.

**CELL BIOLOGY AND GENETICS 1973**

<table>
<thead>
<tr>
<th>AIM</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. to expose a student to a minimal body of factual information which is essential to his future work in this or other parts of this course.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. to integrate a recapitulation of 'A' level work into the new material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. to stimulate a student to read recommended literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. to inform a student of sufficient concepts and terminology to enable him to gain further knowledge from the recommended literature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. to stimulate reading outside the lecture context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. to encourage a student to accept that statements about physics and chemistry may be taken as a starting point for the purposes of constructing biological hypotheses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. to give examples of hypotheses which may be advanced to relate biological phenomena to accepted 'facts' about the properties and behaviour of molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. to give examples of observations and experiments which demonstrate how certain experiments constitute evidence for a particular hypothesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. to show students that biology is a developing science in which new observations and concepts are being made and older ideas modified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. to enable a student to make criticisms and discern gaps in knowledge or explanations, in what he reads</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. to enable a student, (albeit wrongly), to suggest alternative ideas to those found in literature he has read</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. to make a student aware of the diversity and variability of biological phenomena, and of the different types of explanations that may be put forward to interrelate them</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. to lead a student to be aware that descriptions of all structure and function are conceptual models, or hypotheses, that have been suggested by observational or experimental results; so that statements of 'fact' need to be qualified accordingly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. to stimulate the student to attempt to understand new discoveries and interpretations as they occur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. to give a student an interest to investigate the subject further</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MEANS ON A 5-POINT SCALE

For each aim the two blocks represent the difference between ratings on each scale, shown below:

A-B: The importance of aims in the present course
A-C: The importance of aims in the students' ideal course
A-D: The probability of aims being examined
B-D: The degree of success in the present course

The arrows represent the direction of positive difference between the mean ratings on each scale.
and in the present course. However, the course is not very successful in achieving this aim, and this is reinforced by the examination which is perceived to rate this aim very low indeed.

That is the diagnosis of the aims questionnaire, and it can be continued for all the aims; some will exhibit major discrepancies, others will show none at all. The next step is outside the realm of sole responsibility of the evaluator. It is to discuss with the lecturer the implications of these findings for his course, and for him to plan a strategy for the implementation of change either with or without the assistance of the evaluator.

In the present case no changes were made after the first year study, and the course remained apparently unchanged. That is, no formal series of changes was consciously introduced by the lecturer. The following year the aims questionnaire was administered again, and further results were calculated in the same way.

Returning to the example of aim 15, we can see some of the changes that took place in students' ratings, and also examine the discrepancies relative to the 1972-73 ratings. Firstly, the ideal is slightly depressed, though, not significantly so. However, the rating of the importance of this aim in the present course is massively lower than the previous year. The success of the course in achieving the aim and the probability of this aim being examined are depressed by similar amounts which are statistically significant (p < 1%). In this year the major discrepancy is between the students' rating of importance in the present course and the ideal importance. In fact, if one inspects the other scales, one can see this has also happened with ten other aims. For aim 15, the rating of importance is now similar to the rating of success. Taken in isolation this could be an encouraging finding; but when considered in conjunction with the ratings of ideal importance and examinability the opposite is true. Although the detailed
profile has changed from the previous year, some aspects remain the same. They are that students still see this aim as ideally the most important, however it appears to be less manifest in the course; the chance of it being examined is as low as before, and the success of the course in achieving this aim has changed only a little.

This kind of analysis can be repeated for all aims. If this is done, one finds that for both 1971-72 and 1972-73 the following main points arise:

1. Aims concerned with encouraging students to read and to work on their own are always rated very highly by students for an ideal course.

2. These ideals are, not unexpectedly, not manifested so highly in the students' perception of the importance of aims in their actual course.

3. The courses are seen to be relatively unsuccessful in achieving aims concerned with reading and independent study.

4. The examinations are seen to be unlikely to test for aims of this kind.

There are also some differences between the years 1971-72 and 1972-73, the most notable are being substantial upward shift, in the ratings on scale A: aims 3, 4, 5, 10, 11, 12, 13, '14, 15. This can be interpreted as an increase in the student's awareness of the aims of the course, Fig. 5.3.

5.7.2 Pre and Post-attitude Measures

The assumptions underlying the analysis of the results were as follows:
A: Importance of aims in present course
B: Importance of aims in students' ideal course
C: Probability of aims being examined
D: Degree of success in present course
1. If there is a change between the pre-course and post-course ratings then it can be attributed to the effects of the course.

2. Statistically significant changes will be behaviourally significant (2). That is, they will represent an important effect of the course on the students.

3. The concepts and scales validly measure what they apparently measure, namely, given affective dimensions relevant to the course.

   For each scale a mean score was calculated over all students. Assuming that the mean reflects the norm for the population, means were compared for each scale for before the course and after the course. A t-test statistic was computed between the differences of the means using matched samples (52). The means and differences were then displayed in the diagrams shown in Figures 5.4 and 5.5. The direction of the arrow indicates the direction of movement of the mean from before to after. For those differences that were not significant (p > 5\%) a block is drawn at the mid point between the two means. This is done to simplify the diagram to make discussion clearer.

   These changes were then compared to the changes that the course lecturer had hoped for. These are presented in Figures 5.6 and 5.7. They show, for each scale on each concept, a comparison between the lecturer’s expectation of student change during the course and the student changes that were statistically significant (p < 5\%). For 1971-72 they show seventeen changes that were in the direction anticipated by the lecturer on the concept relating to students perception of the course as a whole, and two negative changes. From precise to vague and from orientated towards concepts to orientated towards facts. There were only ten positive matchings on all the other scales and concepts; a number that could have occurred by chance alone (p < 1\%); and thirteen negative matchings (p < 5\%).
The differences of ratings on the scales shown, before and after the course.

Cell Biology and Genetics, 1972

Fig. 5.4

For each concept, the six blocks represent the difference of ratings on each scale shown below, before and after the course.

- A. My expectations towards the course looking back on the course.
- B. Genetics.
- C. Process-chemical explanations of living processes.
- D. The year.
- E. Credits.
- F. Disk.

Arrowheads indicate the direction of change during the course. Non-significant differences are represented by blocks midway between the two sections.
### Fig. 5.5

<table>
<thead>
<tr>
<th>Scale A: What I think the course is going to be like</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale B: The Gene</td>
</tr>
<tr>
<td>Scale C: Cytoplasm</td>
</tr>
<tr>
<td>Scale D: DNA.</td>
</tr>
</tbody>
</table>

#### Semantic Differential: Before and After 1973

<table>
<thead>
<tr>
<th>Scale</th>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Precise** - **Vague**
- **Familiar** - **Unfamiliar**
- **Simple** - **Complex**
- **Easy** - **Difficult**
- **Chaotic** - **Ordered**
- **Meaningful** - **Meaningless**
- **Relevant** - **Irrelevant**
- **Important** - **Unimportant**
- **Rational** - **Intuitive**
- **Interesting** - **Uninteresting**
- **Valuable** - **Worthless**
- **Unnecessary** - **Necessary**
- **Comprehensible** - **Incomprehensible**
- **Satisfactory** - **Unsatisfactory**
- **Systematic** - **Unsystematic**
- **Experimental** - **Non-experimental**
- **Logical** - **Illogical**
- **Pleasing** - **Annoying**
- **Successful** - **Unsuccessful**
- **Structured** - **Structureless**
- **Classical** - **Modern**
- **Oriented towards concepts** - **Oriented towards facts**
- **Never intellectually exciting** - **Always intellectually exciting**
- **Rewarding** - **Unrewarding**
- **Opportunity for originality** - **No opportunity for originality**
- **Never dull** - **Always dull**
- **Factual** - **Non-factual**
- **Flexible** - **Inflexible**
- **Efficient** - **Inefficient**
- **Useful to me** - **Useless to me**
- **Changeless (scales C-D)** - **Changing (scales C-D)**
Although the number of matchings is not statistically significant we may ask whether there are any results that are behaviourally significant. One measure of this could be a systematic change on one scale over many concepts. Here there are two cases. On all three concepts that were included for the scales, there was a negative matching between the lecturer and the students on the scale: orientated towards concepts - orientated towards facts; and a positive matching between them on the scale: never dull - always dull. That is, students experienced a move towards - orientated towards facts when the lecturer had hoped for the opposite; and students experienced a change towards the course being never dull which is what the lecturer had hoped for.

For 1972-73, the results are somewhat different. See Figure 5.5. Comparing these changes with those that the course lecturer had hoped for, Figure 5.7, an entirely different pattern can be seen. There are only two changes in the direction anticipated by the lecturer on the concept of the course as a whole: complex to simple and orientated towards facts to orientated towards principles; whilst there are twenty-three changes in an undesirable direction. For the other scales, slightly shortened in 1972-73, there were seven positive matchings and eight negative matchings ($p<.05$).

Overall these can be interpreted as indicating that the course was much less successful in 1972-73 in changing student attitudes towards the central concepts of the course. The only concept for which there was no obvious deterioration was that of D.N.A. In 1972-73 the number of positive matchings had doubled (from two to four) and the number of negative matchings had decreased (by one). In particular there was a direct reversal of polarity on two scales: Simple/complex, and easy/difficult.

5.7.3 Ancillary Measures

The results for the scale measuring the degree of
w
*X3

H
O
O
P
to
to

>j
p

rt

it

o
I
o

P
O
to

HO P 4
P P
to 3
H*
O O
P Hi P •
to
t-1

F ig . 5.6

o
o
p
H
to
P

-

precise/vague
fami1iar/un fami 1iar
o
+

simple/complex
easy/difficult
o

chaotic/ordered

o
3

O
P
H
H*
CO
O

meaningful/meaningless
superficial/profound

P

important/unimportant

o
Hi

rational/intuitive
interesting/
uninteresting

CO
rt
P

Hi

Hi

valuable/worthless
unnecessary/
necessary
dynainic/static
satisfactory/
unsatisfactory
objective/subj ective
experimental/
nonexperimental

P
P

o
m
r1

p

to

rt

do

P
P

C

P
r+

KH
O

o

CD

CD

<
H*
CD

s:

to

logical/illogical
p1easing/annoying
successful/ ^ ,
unsuccessful
positive/negative
o

+

o
10
P P
H i P.
H i CD
P
P4 rt
P
p. H
P
P rt
O H*
rt P
Oq
P CO
P
rt to
H* P4
O O
to

(o to

rt r+

rt rt

H*
t j CD
P P.

P P
Hi P
H> P
P
P4 rt
P
P H
P
P rt
O H*
rt P
Oq
to

•o p
p

p

rt

rt

CD P
P O

P p
P.
o
O p4
P p
P P
Cq
P
S»

O o
P p4
CD p
p
Cq
CD
V
P
P
P.

O'
P

rt

O

rt rt

to

to

rt

rt rt

oP
P P ►
Hi P o P
to P
Hi P
P H- P
t r rt

rt rt

P
P

P

H
P

P

rt

rt

P

P

H*

H* Cq

O 10
H*
P
rt to ►
O to
H* P 4 P P4
O O rt O
H* s:
P ?
P

+

+

CO to

P P
P

O
P
P

P

O
o

p4
p
p
Cq
P
V*
O'
P

rt

+
+

I

rt

O

H

P

rt
H*

s: P
P 'C q
P to

rt

to
to p4

rt

O

P S!
Hi P
Hi P

P4P
P
P O
p4
P p
P p
rt Cq
H* p
CD V
|-i.
o4
P p

rt rt
P

to

P P
Hi P
Hi P
P
P rt
P
rt H
H- P

O rt
H* H*
•o P
Q
P C
rt to
P
P to
P

P4
O
£

to P
H* P
3
H* P
1
—1
P O
H P4
P
O P
P Cq
P P
W
P
P
P

o
p
o

o
p
o
CD

fa

rt
tn

classical/modern
H
orientated to1 concepts/ (D
t—•
orientated to’ facts
P
rt
never intellectually
Hexciting/always....
P
rewarding/unrewarding
opportunity for
onginality/no.,
never dull/always dull
factual/nonfactual
real/imaginary

Cq

rt

O
rt

P4
CD

O
O
p
H
to

CD

efficient/inefficient
useful to me/
useless to me


<table>
<thead>
<tr>
<th></th>
<th>'whole course'</th>
<th>The Gene</th>
<th>Cytoplasm</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>flexible/inflexible</td>
<td>+</td>
<td>++</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>precise/vague</td>
<td>+</td>
<td>-</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>familiar/unfamiliar</td>
<td>+</td>
<td>-</td>
<td>0+</td>
<td>++</td>
</tr>
<tr>
<td>simple/complex</td>
<td>++</td>
<td>-</td>
<td>0+</td>
<td>++</td>
</tr>
<tr>
<td>easy/difficult</td>
<td>+</td>
<td>0+</td>
<td>0+</td>
<td>++</td>
</tr>
<tr>
<td>chaotic/ordered</td>
<td>+</td>
<td>0+</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>meaningful/meaningless</td>
<td>+</td>
<td>-</td>
<td>0+</td>
<td>-</td>
</tr>
<tr>
<td>orientated to concepts/orientated to facts</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>important/unimportant</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>comprehensible/incomprehensible</td>
<td>+</td>
<td>0+</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>systematic/unsystematic</td>
<td>+</td>
<td>-</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>structured/structureless</td>
<td>+</td>
<td>0+</td>
<td>0+</td>
<td>0+</td>
</tr>
<tr>
<td>rational/intuitive</td>
<td>0+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interesting/uninteresting</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>valuable/worthless</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unnecessary/necessary</td>
<td>0+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>satisfactory/unsatisfactory</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>experimental/non-experimental</td>
<td>0+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logical/illogical</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pleasing/annoying</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>successful/unsuccessful</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>classical/modern</td>
<td>0+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relevant/irrelevant</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>never intellectually exciting/always intellectually exciting</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rewarding/unrewarding</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>opportunity for originality/no opportunity for originality</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>never dull/always dull</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>factual/non-factual</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficient/inefficient</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>useful to me/useless to me</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**KEY**

++ student ratings showed a change, and staff anticipated a similar one.

+ student ratings showed a change, but opposite to what staff had anticipated.

0+ student ratings showed no change, but staff anticipated one.

+0 student ratings showed a change, but staff had not anticipated one.

student ratings showed no change, and staff had not anticipated one.

*Fig. 5.7*
understanding of course content are given in Figure 5.8. Notable features are:

1. the very low rating of the Meselson-Stahl experiment.

2. the overall relatively low ratings of topics from gene-enzyme relationships onwards. That is, most of the genetics section of the course.

If one inspects the open-ended question asking for comments on aspects of the course one finds the following replies:

1. 32 students did not respond.

2. 15 students make specific criticisms of the Genetics section of the course using statements such as: boring, confusing, vague, not systematic enough.

3. 10 comment that the course was generally vague.

4. 9 mention repetition of 'A' level work.

5. 7 found the course uninspiring or boring.

6. 6 asked for the course to include more stimulation to read outside lecture content or asked for specific references to books to be made.

For the question asking for reactions to filling in the questionnaire, the following replies were given:

1. 18 said it was difficult to understand or the wording was rather complicated.

2. 15 gave no response.

3. 15 found it tedious or boring.
Fig. 5.8 Degree of understanding of Course Content

1: Structure of protein molecules
2: Electron microscopy of cells
3: Composition of the cell membrane
4: Active transport
5: Function of cell organelles
6: The endoplasmic reticulum
7: Krebs cycle and oxidative phosphorylation
8: Chromosomes
9: DNA structure and replication
10: The Meselson-Stahl experiment
11: Mitosis and meiosis
12: The Mendelian theory of particulate genetics
13: Gene-enzyme relationships
14: Crossing-over and genetic recombination
15: Defining the gene
16: Chemical basis of mutation
17: The cistron and cis-trans test
18: Genetic code and protein synthesis
19: Bacterial conjugation
20: Genetics of viruses
21: Regulator and operator genes
4. 13 said it was repetitive.
5. 11 mentioned that they had difficulty in deciding what to put.
6. 7 thought it too long.
7. 7 commented that it was worthwhile if it helped the course.

In addition it seems appropriate to add an observation of the students' reaction to being asked to complete the questionnaire. In 1971-72 the investigator explained the nature of the research project and the relevance of the questions that were to be asked to the group of students before administering the questionnaire. No questions were asked or comments made either at that time, during, or following the questionnaire. In 1972-73 the same procedure was followed under essentially the same conditions. However, the response was strikingly different. Many questions were asked by students, comments were made during the session about difficulties in understanding words or sentences in the questionnaire, and other comments were made regarding some students ambivalent attitude towards the course. It must be emphasized that the total number of students making comments openly was very small (~5%). Nevertheless there was an apparently significant change in the reaction of the class. This may be spurious, due to factors outside the control or awareness of the investigator, but such factors are present for all research on students using a group administered questionnaire.

5.8 Discussion of Results

In the first year of the study (1971-72) an attempt was made to use two different measures of course effect: the aims questionnaire, and the attitude measures. These were designed to measure two different aspects of the course and as such it
is impossible to compare them directly. However, in the following year ancillary measures were introduced to enrich the detail in the findings and to provide a cross-check on some of the outcomes from the two main measures. These do not negate the results of the chief measures, and they give support to some of them:

1. The aims questionnaire showed that most students thought that reading outside the lecture content was a very important activity which the course had not sufficiently encouraged. Six students support this finding in their open-ended comments.

2. The attitude measure indicated that student views of the gene had become more vague, more meaningless, more unimportant, more structureless (p all < 1%). The index of understanding scale showed that the parts of the course dealing particularly with the gene were understood less well than the rest of the course. The adverse comments on the open-ended question were directed particularly at the genetics part of the course. These findings are consistent. However, there was also a change in the attitude scale on the gene towards more simplicity. This is not consistent with the other findings.

There is another problem. There are differences between the 1971-72 results and the 1972-73 results. Overall the 1972-73 results are much more critical of the course, and in particular, the attitude scales show many negative changes from the first to the second and in some cases direct reversals of opinion.

All the measures indicate that there were some major differences between the two years and that these suggest a deterioration in the course. However, there is nothing contained in the data that was collected that pinpoints any substantive reason for this. Neither the evaluator nor the lecturer could suggest any explanation based on the evidence,
both formal or informal, that accounts for such a change. Indeed, the results from the various scales surprised both the evaluator and the lecturer.

None of the obvious explanations that could be suggested have any support in evidence: it was possible that the lecturer had been made self-conscious in the first year and that this had affected his teaching adversely in the following year; it was possible that there were differences between the two student groups. However, none of the evidence that was at our disposal supported these interpretations. In particular, neither 'A' level grades or first year examination results were significantly different from one year to the next.

The fact that the results remain problematic does indicate a limitation of the evaluation methodology that was adopted.

5.9 Outcomes

Throughout the present study the investigator was in close touch with the course lecturer; results and analyses were discussed as soon as they were prepared, and the lecturer contributed many ideas towards the content of the measures and the administration of the instruments. At no time was any antagonism shown towards the investigator or his work. Nevertheless, the study failed to produce any substantive change in the way that the course was organized, the content of the course, the aims of the lecturer or the type of examination.

Some of the possible changes were not particularly allowed for in the design of the research. The teaching method was given and fixed by external constraints, the overall subject areas were similarly prescribed, the aims of the lecturer provided the framework for the study and the style of examination was to a great extent externally imposed. Any findings that suggested a change in those areas that were externally fixed could not be acted upon. Many findings were of this type.
For example, the low examinability of some of the aims, and the low degree of success of the course in achieving them, particularly in those areas dealing with encouraging and enabling students to work independently and to read outside of the course; pointed directly to changes in the teaching method: away from a didactic style, and to changes in the examination: towards specifically providing and encouraging answers away from subjects covered exclusively in lectures. Of course, it is possible to modify the styles of lecturing and examining slightly to allow for changes in the areas suggested, but it was considered that the effects of this would be marginal.

Another unexpected outcome was the difference between the two years of the course, which were ostensibly the same. There appears to have been a change for the worse in the course. There were indications both in the aims questionnaire, where there was an overall decrease in ratings of the importance of aims in the present course; and in the attitude measures, where there were many more negative changes in 1972-73 even allowing for the differences in pre-test scores. However, no apparent change in a course can, in itself, be interpreted as an actual negative change in some circumstances. If there has been a regular modification to the course over a period of years, then when this change stops, the rate of change decreases. Stagnation of a course is in itself normally a negative effect.

However, the present study was not originally conceived as a planned strategy for course changes. It was an exercise to determine whether useful information can be assembled on a course, within rigid constraints, that could be used to help decision-making concerning changes to the course. Of course, the validity of such a method can ultimately only be tested by examining the quality of resultant decision-making. If this does not improve, then the methods of evaluation are not useful and should be dropped.
Also, it is necessary to be able to account for the differences between the two years of the course empirically. If it is not possible to find a substantive cause for the effects that were measured, then the entire evaluation methodology is suspect.

Finally, what has been the effect on the lecturer's attitude towards the course as a result of engaging in this study. An open-ended interview with the lecturer was arranged. The following statement, taken from this summarizes the lecturer's view of the course at that time:

"I think overall, that the aims as we have them for this lecture course are really too ambitious for students at this stage of the course. I'm beginning to doubt whether it is worth trying to do that sort of thing".

When asked if he would want to change the aims and make them a little less ambitious his reaction was that he would sooner maintain the level of the course rather than, perhaps, achieve a lesser set of aims.

5.10 Statistical Footnote

5.10.1 Reliability

The aims questionnaire measure was tested in a similar manner to that described for the instrument in section 4.10.1 to obtain an estimate of the reliability of the scale. A split-half test was performed over all four scales: A to D, for both the 1971-72 and 1972-73 results. The values thus obtained for the product-moment correlation coefficient, \( r \), and the Spearman-Brown split-half reliability coefficient, \( R \), are displayed in Figure 5.9. As all the coefficients obtained were high then it was assumed that the aims questionnaire scales did measure some property in a reliable way.
5.10.2 Validity

It was not possible to apply the same test of face validity to the biology aims questionnaire items as described in section 4.10.2. The aims were established, in this case, by one particular lecturer for one particular course. A test could have been made to see whether or not other lecturers thought the aims appropriate for a course of this type and level, but this would not have been in keeping with the aims of the study which were to develop an instrument suitable for the given lecturer in the given course. In one sense the items were valid a-priori as they had been constructed for the sole purpose for which they were used. They were, however, checked to see if they were understandable. Two non-biologist members of staff and three biology students were presented with the questionnaire before it was administered and then discussed its clarity with the evaluator. The suggestions were incorporated with the final instrument.

5.11 Conclusion

The study met its declared aim of providing information pertinent to course improvements. However, it was not able to test this empirically by resort to examining observable changes in decision-making behaviour in the course. Two outstanding problems remain. Firstly, there are the unexplained differences found between the 1971-72 course and the 1972-73 course. Secondly, there is the problem of influencing decision-making in such a way that action is taken as a result of evaluation. It is believed that from the start of the study a relationship inappropriate to the immediate needs and interests of the lecturer was established. It is also believed that if a different relationship to that of researcher-researched had been initiated then the evaluation would have been different, but would have been more easily implemented. The evaluation that would result may have been less formal and analytical, but could be no less productive.
<table>
<thead>
<tr>
<th>Scale</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>r</td>
<td>R</td>
<td>r</td>
<td>R</td>
</tr>
<tr>
<td>Sample 1</td>
<td>0.78</td>
<td>0.88</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>Sample 2</td>
<td>0.87</td>
<td>0.93</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>Sample 3</td>
<td>0.95</td>
<td>0.98</td>
<td>0.93</td>
<td>0.96</td>
</tr>
</tbody>
</table>

1972

<table>
<thead>
<tr>
<th>Scale</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>r</td>
<td>R</td>
<td>r</td>
<td>R</td>
</tr>
<tr>
<td>Sample 1</td>
<td>.97</td>
<td>.99</td>
<td>.95</td>
<td>.97</td>
</tr>
<tr>
<td>Sample 2</td>
<td>.89</td>
<td>.94</td>
<td>.90</td>
<td>.95</td>
</tr>
<tr>
<td>Sample 3</td>
<td>.95</td>
<td>.97</td>
<td>.97</td>
<td>.98</td>
</tr>
</tbody>
</table>

1973

Biology aims questionnaire: Correlation and reliability coefficients for random split-half tests: 1972 and 1973

FIGURE 5.9
It is appropriate to ask in conclusion 'what has been learned from this experience'? An indication of some of the outcomes for the lecturer has already been expressed. The researcher sees it somewhat differently. In the strict terms of the original research aim the research has been successful. However it has succeeded at the expense of sensitivity to the real situation in the course, and at the expense of usefulness to the immediate problems of the course. What is meant by this is that the researcher believes that some of the most fundamental issues and problems concerned with university courses have been systematically ignored. These issues are concerned with the purpose of the course and the commitment to this purpose by lecturers and by students, and with the methods used to achieve this purpose. To examine a course from the viewpoint of predetermined aims and objectives and of fixed method: the lecture is to make the assumption that these are the best available to the lecturer and to the students, and that both groups have agreed that this framework is valid. This assumption has not been shown to be the case. And no attempt was made to do this. This is a limitation to the research and should be used to place it in context. It is one example of one strategy of formative course evaluation. Further studies would need to commence by demonstrating the validity, or otherwise of this assumption. If that cannot be done, then they will be of as limited use as this one.

In this study results emerged that were not explainable within the framework that was established. It is the belief of the author that the limitations are not to do with the technical limitations of the evaluation instruments that were used but were intimately connected with the evaluation strategy that was set up. If the results from the evaluation instruments come as a complete surprise to both the teacher and the evaluator then a substantial area for investigation is pinpointed. That this area was not pursued in depth is one limitation of this case study.
6. CASE STUDY: QUANTUM MECHANICS AND A NEW TEACHING METHOD

6.1 Introduction

This chapter presents the evaluation of a course that was highly innovative. The course was in Quantum Mechanics taken by second year students studying for degrees in Physics and Physical Sciences. Lectures were drastically reduced and students worked on a scheme of self-paced instruction known as the Keller Plan (67). The evaluation of the course was stimulated by the instructor's need to investigate this innovation to determine the extent to which it was proving successful.

The main aim of the evaluation was to monitor the course to determine its strengths and weaknesses and to feed back information about areas for improvement to the course organizer so that the course could be adapted to the needs of the subject and to the needs of the students who were taking the course. This is the aim of the formative evaluation activities that took place.

In addition to this fairly short-term study an attempt was made to lay the ground work for an investigation of the Keller Plan system as a method of teaching the subject Quantum Mechanics. Information in this area would be used to aid decisions about the future methods of running this particular course and of courses in other areas. A subsidiary aim of the study was to derive information that could be used to assess the teaching method. Much of this is only preliminary, but an indication of the direction of this work will be given.

6.2 The Course

The second year course in Quantum Mechanics was designed for students who had taken introductory courses in Atomic Physics. The study reported in detail here took place during the first two terms of the 1971-72 academic year. The course
began at the start of the Autumn term and officially finished mid-way through the Spring term. It was assessed by one of the Part I examination papers in mid-June. Seventeen students studying for an honours physics degree were enrolled with thirteen students on an honours degree in physical sciences. The physics students had been selected as being in general of lower ability than the full intake for the physics degree. They formed the B group which consisted of students who had performed less well than the A' group in a range of examinations taken in the previous Easter vacation. The physical science students comprised the entire intake for the Physical Science degree course.

The course was arranged in a series of fifteen units. Each unit consisted of written material including:–

(a) a statement of the objectives of the unit in terms of what a student should be able to do on completing it,

(b) a statement of previous knowledge that was required of a student before tackling the unit together with an indication of where that pre-knowledge could be acquired (references to books, notes, etc.),

(c) a reference to specific pages and sections of a textbook that the student should study,

(d) some notes that clarified or expanded the explanation of the topic found in the textbook,

(e) problems that students could attempt to see if they had understood the unit,

(f) the solutions to the problems including the derivation or calculation in detail.

Students studied these units on their own in their own time. When they felt that they had understood the unit they were asked to present themselves for a short test, typically taking them 20 minutes, which examined their achievement of the unit objectives. When they had finished the test
students immediately discussed their results with a tutor. If they satisfied the tutor that they had mastered the test then they were allowed to proceed to the next unit. For administrative reasons students had one unit in hand at any time. This enabled them to continue their study if there was not a convenient time for them to be tested. Testing took place in the periods that had previously been occupied by lectures and tutorials. There were two one-hour testing sessions per week in the first term and three one-hour sessions per week in the second.

Students were able to discuss their problems with the tutors in the testing period without taking a test. This usually occurred in the first twenty minutes when tutors were not occupied with testing. In addition, four lectures were given throughout the course. Three of these were stimulus lectures and aimed to interest students in topics which were related to the subject of the course, but which were not in the syllabus and were not therefore examinable. The other lecture was given near the beginning of the second term and was designed to give students a broad review of the course to show them how the individual parts were interrelated. At the beginning of the course students were introduced to the teaching method by way of an introductory talk by the course organizer.

In this way the conventional 35 hour lecture course was replaced by a self-paced course of 15 units. The aims and objectives in terms of the subject matter to be studied remained the same as before but the aims that were related to the method of study were radically different. It was expected that students would be encouraged to work on their own and become more independent in their study habits; that they would all achieve mastery through the lack of time constraints; that they would be able to study at times prescribed by themselves without the lock-step of the lectures; and that they would have more opportunity than before to discuss their problems and difficulties with the tutors.
The course was, as far as the designers were aware, the first course to be run in this way in Great Britain. It was regarded as very important by the course designer that the course be evaluated and monitored intensively by an evaluator. In order to reduce the possibility of subject matter problems developing two additional staff who were subject experts acted as tutors for the first term during all testing sessions. The course was therefore highly untypical. It was the subject of careful scrutiny both by an evaluator and by an additional teaching contingent.

A flow diagram of the course design is given in Fig. 6.0. Lectures are indicated at the points that they would be received by students who were proceeding at the recommended pace. That is, the pace at which they would finish the course in 15 weeks.

6.3 Background to the Course

The scheme for the course outlined in the previous section resulted from two influences. Firstly, the course had been subject in the 1970-71 session to substantial innovations and an evaluation of these had pointed to certain problem areas both in content and method. Secondly, a self-paced teaching method known as the Keller Plan had come to the attention of the course organizer whilst he was visiting the Massachusetts Institute of Technology in the same year.

6.3.1 The earlier Course

During the session 1969-70 when the Quantum Mechanics course was given for the first time, it was given entirely through lectures and subject tutorials. The lecturer was very dissatisfied with the reception that the course was getting from students: they lacked interest, very few attended tutorials and the examination results were disappointing to the lecturer. A group of three
Fig. 6.0

Pass test; go on to next unit of material.

Fail test; study the unit of material again.
educational technologists including the present evaluator were invited to look at the course and make recommendations which they thought would help improve the course. They sat in on a few lectures and tutorials and talked to a few students. They produced a set of recommendations for possible action. Some of these were outside the range of influence of the lecturer and were not implemented; others were implemented in full.

The outcome was a greatly changed course. Lectures remained the central feature, but tutorials were replaced by examples classes. These took the form of the total class meeting in one room with the three tutors rather than separate small groups with a single tutor. Problems were provided which students were encouraged to attempt prior to the class. Difficulties in these were then discussed in the examples class either on an individual one-to-one basis with a tutor, or if demand was sufficient a problem would be discussed on the blackboard by one of the tutors. This, in itself, was not a major innovation. It had been used successfully in teaching mathematics to the same group of students. In addition, however, for each section of the course students were provided with the aim of that section, a list of specific objectives in detail, a list of the pre-knowledge required for each objective and a flow diagram illustrating the interrelation of different parts of the section. This handout was given to students prior to the lectures that covered that part of the course. Also, at the end of each examples class, students were given a short, ten minute test with multiple-choice items. These were displayed on an overhead projector transparency and students attempted them on the spot. They were immediately given the correct responses and they recorded their performance on each question on a sheet which gave references for students to follow up on each question that they had answered incorrectly.
The course given in the 1970-71 session was subject to a detailed evaluation which investigated each element of the teaching system that was used. Details of the findings appear in an internal report. The two main points were as follows:

1. "By far the most successful innovation was the introduction of the short tests given at the end of examples classes."

2. "The greatest problem with the pre-lecture handouts was their lack of use."

The report concluded with the following statement:

"In general terms the effects of the course in terms of student achievement have come much closer to the expectations of the lecturer than when the course was given in the traditional manner. Nevertheless the results of the evaluation do suggest certain areas where the course is under-achieving. These areas are not those of student performance, or more accurately student performance is probably not the greatest of these. They are in the area of the transactions which take place during the course between staff and students. As mentioned in the preliminary evaluation, "the main problem is one of getting students 'involved' in the course." A decision needs to be made not simply on whether certain innovations need to be retained, improved or rejected, but whether there should be new innovations. It is possible that greater student achievement can be obtained by an extension of the present methods, but it is the opinion of the evaluator that a greater course improvement can result from innovations which affect the transactions of the students with the subject matter, and with the lecturer and the tutors."

6.3.2 Influence of Keller Plan Courses

In 1968 a paper by Fred S. Keller entitled "Goodbye,
"Teacher..." was published (68). This described a method which Keller and his associates had started to develop in 1962 to teach psychology to undergraduate students at the University of Brasilia. He distinguished the features of this method from conventional teaching procedures as follows:

1. "The go-at-your-own-pace feature, which permits a student to move through the course at a speed commensurate with his ability and other demands upon his time.

2. The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of that which preceded.

3. The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information.

4. The related stress upon the written word in teacher-student communication, and finally

5. The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process."

The teaching method based upon these principles was known as a Personalized System of Instruction (PSI) or the Keller Plan. This method received great attention from university and college teachers throughout the United States, see for example (64,69,79,83). and it was used in the teaching of undergraduate physics at MIT by Ben A. Green. Reports of this work (57) attracted attention of the community of physics teachers in higher education and the Keller Plan became to be increasingly used in the teaching of physics as well as psychology, and has spread to many other subjects (117).
The most important feature which was adapted from the Keller Plan for the Quantum Mechanics course at Surrey was that of self-pacing and it was this aspect that made necessary the removal of regular lectures from the course. If students were pursuing the course at a rate that was appropriate to them individually then it would not be possible to introduce a talk at a fixed point in time and expect many students to benefit. In addition the Keller Plan introduced the feature of unavoidable staff-student contact. This was expected to effect the transactions of the students with teaching staff which were unsatisfactory in the previous course.

The Keller method provides one means of tackling some of the problems that the previous course evaluation had exposed. In particular it was possible to retain the component of frequent testing whilst at the same time changing the transactions between students and staff. Staff time during the course could be devoted to the individual testing of students performance and the discussion of student difficulties, rather than to the development of a performance on the part of the course organizer in a lecture. The favourable elements of the previous innovation could be retained and a new component could be added if the Keller Plan was adopted.

6.4 Specific Constraints

The context in which the course was situated produced constraints on the details of the teaching method which was finally selected and on the kinds of evaluation which could be attempted.

6.4.1 Constraints on the Method

The major constraint was that imposed by the necessity for the course to fit into administrative arrangements that
were centrally governed. These prescribed the times of
meeting and the place of meeting. They also provided a
constraint which was considered likely to affect the
success of the scheme. This was the need for the course
to be examined in a conventional three-hour examination
paper ten weeks after the end of the course. The
arrangements were slightly different for the two degree
courses involved, but in both a single paper included
questions from two subjects one of which was Quantum
Mechanics. Keller Plan courses in the United States
usually assessed students in terms of either the number of
units covered, or they awarded an automatic grade A for
completion of all units. (57). In general little or no
weight was carried by an additional test or examination.*
However, in the present context, the examination was the
only officially recognised form of student assessment.

A common element of other Keller Plan courses was
the use of undergraduate student proctors as tutors who
marked unit tests and discussed them with students. These
proctors either received payment for their services or they
were awarded credits which could be counted towards a degree.
In the British context neither of these opportunities are
available: Students on grants cannot officially work during
term time and no university awards credits for teaching
(at least not to non-education students). Tutors are either
teaching staff or postgraduate students working on a part-
time basis. They tend therefore to be more expert, but
also more remote from undergraduates. The Keller Plan had
proved successful with student proctors - it was assumed
that now it would work with academic staff as tutors.

6.4.1 Constraints on the Evaluation

The present evaluation was governed less by external
factors than the other case studies, and more by problems

*Currently (1974) many courses in the U.S. given by Keller
Plan now give weight to a final exam. (157).
intrinsic to the method. Some of these were minor. For example, except at the beginning of the course students were not present in one place as a group, at any given time very few were present. Others were more substantial; one particular constraint on the evaluation was the external requirements of the examination. The examination paper could not be effectively used as a measure of student achievement on the course. One problem in this was that each department had a separate procedure. In one the paper was divided equally between two subjects and students were required to answer a given number of questions on Quantum Mechanics, in the other students were required only to answer five questions out of eight, of which three were on Quantum Mechanics and five on subjects not requiring any knowledge from the Quantum Mechanics course. In the latter case, if students were prepared to risk not having a choice of questions to answer they could effectively drop Quantum Mechanics without penalty. This great flexibility did mean that very little meaningful information could be derived from examination marks. It was considered highly unlikely that students would be prepared to sit any other tests in the subject so that no independent measure of student achievement was available with which to compare the progress of students on the units.

The role of the evaluator was somewhat different in this case study also. The course organizer was committed to innovation and to the value of evaluation before the evaluation exercise was initiated. He had initiated the evaluation of the course and had invited the collaboration of the evaluator. In addition a working relationship between the innovator - the course organizer - had been established previously; they both worked in the same unit and had collaborated on other course innovation and evaluation projects.*

*Most importantly, the course organizer and innovator was the supervisor of the evaluator for work towards a higher degree. This was not seen by either party as a substantive constraint on the task.
All these factors were a positive influence towards removing constraints on the evaluation that might otherwise have caused difficulties. There was a constraining effect though in this relationship when one comes to consider the transferability of evaluation methodology. It may be that some of the evaluation activities of this study could not be attempted in the same way under a different relationship between innovator and evaluator. This problem will be discussed in greater detail in a following chapter.

6.5 Rationale for Choice

The study of the Quantum Mechanics course is set in a different context to those reported earlier. It is an evaluation that was requested by a university teacher who was very interested in and committed to evaluative activities. This is the most important reason why this study is reported here. It was hoped that one outcome of this study would be to help illuminate the differences in the evaluation of university courses that would be possible in situations where the clients of the evaluation were committed to its implementation.

Another factor in the selection of this particular course for evaluation was that it was highly innovative. There seemed a-priori to be a very powerful reason for incorporating evaluation into the course; the course was being given by a method that the organizer had no experience of and this made it important that some form of course monitoring take place to check that all was going well. The same argument is not applicable in practice to the courses given by conventional teaching methods. In these cases the same staff would be fairly confident that they could monitor the course themselves, without the aid of additional resources and it would be difficult for the evaluator to challenge this assumption.
Three aspects of the problem were covered by the evaluation. Firstly, there was a need to monitor students progress on the course, to find difficulties and problems that they were experiencing and to provide information for the course organizer so that small changes could be made to improve the course whilst it was in progress. This was essentially short-term formative evaluation. Secondly, information was gathered relating to the structure of the course and to the students experience of the organisation and the methods that were to be employed. This was to be used to facilitate decision-making about the style and structure of the course for future years. It was anticipated that this aspect might suggest substantive alterations to procedures whilst keeping within the overall philosophy of self-paced learning. This is formative evaluation for medium and long range decision-making. Finally, a start was made to prepare a basis for a longer term summative evaluation of Keller Plan and other self-paced courses. Essentially it was expected that information derived under the first two headings would be useful for this. This was a distinct aim of the total evaluation and will be discussed as such although little attempt was made to incorporate special procedures in this area in addition to those selected for formative purposes.

6.6.1 Short-term formative evaluation procedures

Much of the information gathered to provide short-term feedback to the course organizer whilst the course was in progress was informal. The evaluator attended most (~75%) of the testing sessions and the lectures that were given. He observed the number of students present and the activities they were engaged in; after testing he talked informally to students and asked them about difficulties they were having and their ideas and attitudes towards the course. Student difficulties were recorded and reported to the course organizer. The sources of information were kept confidential.
The tutors gained much useful information through the post-testing discussions with students. They reported that this was a very valuable source of information about student learning difficulties and about ambiguities and errors in the units and the textbooks. This was regularly reported to the course organizer and it did not require the support of the evaluator.

One systematic source of information was provided by the unit feedback sheets. These were single page forms that students were given with each unit they received. They were asked to complete them and return them when they collected the following unit. Questions about the difficulty and interest of the parts of the unit were asked together with the time taken to complete the unit, pre-knowledge that was assumed and comments on the text and the problems were solicited. The form is shown in Appendix III. Records of student responses were kept and were also regularly shown to the course organizer.

Feedback on students that were slow or did not present themselves for testing was obtained from the records of student progress that were made. These recorded the date on which each student took each test and the result of the test. The course organizer noted students that were slow and wrote to them to encourage them to see him to discuss their difficulties with the course.

6.6.2 Long-term formative evaluation procedures

1. The informal interviews with students after testing periods also provided a source of questions and problems concerning the effect of the course. Many of these ideas were incorporated into a questionnaire which was administered to students in the eighth week of the course. It was given personally to all students who were present for testing during that week and posted to students not present. They were assured of anonymity and were asked to complete the questionnaire.
and return it to the evaluator personally or through the internal post. The questionnaire aimed to probe student opinion about many aspects of the course including the following (the complete questionnaire is shown in Appendix IV.

(i) Each component of the course, e.g. objectives, problems, feedback sheets, the textbooks etc. These were rated on a scale of usefulness from 1 (very useful) to 5 (not very useful) and students were asked to give reasons for rating it in a particular way. Problems and tests were also rated on a scale of difficulty from 1 (too difficult) to 5 (too easy).

(ii) How students worked on the course, e.g. did they work in class when not being tested?, how much time each week they spent on the course, whether it took up more time than other subjects?

(iii) Students overall impression of the method, e.g. did they think they learnt as much by this method compared to lecture courses, was their time efficiently used, did they feel there was greater or lesser contact with staff, advantages and disadvantages of this method?

(iv) Reaction to given statements about the course. These were taken from actual student comments and students were asked to agree or disagree to the statements. These included: 'The tutors expect too much of us', 'I couldn't work up much enthusiasm for Quantum Mechanics no matter how well it was presented.'

(v) A list of each topic in the course was provided and students were asked to rate each topic in terms of their degree of understanding of it from 1 (don't understand at all) to 5 (understand very well). Students were asked to indicate the unit that they had just completed.
2. After the end of the course just before students were to sit their examinations a second questionnaire was posted to them. This was of an open-ended nature and it asked students to record their overall impression of the Quantum Mechanics course including whether or not they thought that the self-paced method encouraged any feelings of competitiveness or co-operation amongst students. The second part of the question was included to find out about an issue of student competitiveness where the course had been criticised by staff in other departments. A follow up questionnaire was also sent due to the initial low response.

3. The student progress record was also a useful source of information concerning the differential progress of students especially when considered together with the unit feedback sheets and scale of understanding of course content.

4. An attitude scale similar to that used in the Biology course was administered at the beginning of the course and at a time when the course had officially finished. (For details of the construction and use of this see section 5.6.2 Students in the latter session were asked to attend together to complete this schedule and to discuss their experiences of the course. The concepts rated by students were Quantum Mechanics, Mathematics, My expectations towards the course/ Looking back on the course, Wave-particle duality, The uncertainty principle, the wave equation, Probability and Measurement. The first three were rated on the same 30 scales and the rest were rated on a shorter number (20) of these scales.

5. The two tutors and the course organizer, who also acted as a tutor were interviewed at the end of the course. These were semi-structured and aimed to explore the experience of the tutors in working in the self-paced teaching system, their conclusions about its efficacy in the present context and suggestions for possible future changes.
6. Finally the examination results were studied, and as far as was possible, compared to the results obtained in the previous year.

6.6.3 The beginning of a summative evaluation

As mentioned before no attempt was made to set up a formal summative evaluation of the teaching system. However, the information gathered for the formative evaluation was examined and collated to form hypotheses that could be used as a framework for the study of other courses of this type. The intention of this was to provide the initial stages of a framework that could be used as a perspective through which to view self-paced courses. The main sources for this were firstly, students who had participated in a self-paced course, staff who had operated within such a structure, critics of such courses, and finally other proponents of similar schemes. Information from the last two groups was derived from the literature and from reactions of other university teachers to reports that were prepared to describe the methods of the present course and the findings of the evaluation.

An important distinction to be made at this stage is that between the evaluation of a particular system in terms of its structure, and in terms of its content. For example, in the present course it may be found that students were not appreciating the content of the course, or they did not understand the subject. This could be due to one of two factors. Firstly, the course may be structured in such a way that organizationally students had difficulties in learning, i.e. the way the learning system was arranged was at fault. Or, that the subject matter was badly arranged and that with a different subject-matter content the structure of the method would be satisfactory. There are therefore two considerations to be made when discussing the findings of the evaluation: whether they can be accounted for by the structure of the course, or whether they derive mainly from the content of the subject. It was expected that, at least partly, some significant problems would be found in the content part of the course. The course
was being tried in a written rather than a verbal format – there were bound to be mistakes and these could be relatively easily detected. However, the problem remained of how to separate these effects from the effects of the organization of the course – the teaching method. To some extent these problems were never resolved: one example of a method is insufficient evidence from which to deduce information about it.

6.7 Results and Analysis

The individual procedures were intended to interrelate and produce results which could be viewed as a whole entity. Some overlap and repetition was present as a check on the reliability of individual measurements and no one procedure was used to describe any particular outcome of the course. Findings which were acted upon were chiefly those where supporting evidence came from more than one source.

The rationale for the forms of analysis used was that a composite measure based on information from a variety of sources and a variety of procedures is more reliable and valid than information from any one of the single sources or procedures. This is the concept of triangulation referred to by Webb, et al (140).

The results will be presented and discussed under headings which refer to the kinds of evaluative information that they primarily relate to. This is a somewhat artificial division and is more relevant to a theoretical discussion of evaluation methodology than to the dynamic of the decision-making situation. They do, however, roughly correspond to the time sequence of decisions that were to be made. Revision and improvement of the present course for the present students was the most immediate priority. This was followed by the improvement of the course for the next group of students in the subsequent years. The least immediate, but perhaps the most important to non users, was the development and improvement of the teaching method used for the course. This progression
also leads in a sequence from the most tangible to the least tangible. The kind of information required in the first category is more certain to determine, that necessary to satisfy the third category is difficult to deduce and uncertain in its determination. The audience for the first is known: the audience for the last is generally unknown.

6.7.1 Short-term formative results

These were to be used to give week by week feedback to the course organizer on the progress and opinions of students as they studied the course from unit to unit.

Comments from these feedback sheets were recorded and given to the course organizer every three to four weeks (more frequently at the start of the course). These recorded student comments in detail. They mostly took the form of specific difficulties, for example, "the maths took more working out than the physics," "Eisberg's notation is difficult to follow", "I didn't know the Wilson-Sommerfeld quantization rule". It was not possible to obtain detailed figures on the ratings of difficulty and interest as students grouped the parts of each unit in different ways. However, a crude measure of overall difficulty of each unit could be obtained by the average time taken for students to complete each unit and the range of times that were exhibited. See table

One of the greatest problems with the feedback sheets was the low rate of return. The maximum returned for any units were 17 each for the first two units, i.e. approximately one half of the maximum possible return. The sheets did not form an integral part of the teaching system and their return was not required. The students returning them were atypical of the class: sheets were returned by predominantly the students who passed units quickly. This would tend to indicate that the average time taken per unit is a considerable underestimate of the true time taken, although it is possible
### QUANTUM MECHANICS FEEDBACK SHEETS

<table>
<thead>
<tr>
<th>UNIT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO. of STUDENTS</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>11</td>
<td>13</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AV. TIME SPENT (hrs)</td>
<td>2.6</td>
<td>3.5</td>
<td>3.1</td>
<td>3.0</td>
<td>4.4</td>
<td>3.4</td>
<td>3.6</td>
<td>3.9</td>
<td>2.5</td>
<td>3.7</td>
<td>4.7</td>
<td>5.6</td>
<td>2.3</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>RANGE (hrs)</td>
<td>1-6</td>
<td>1-12</td>
<td>1.5-6</td>
<td>1.5-5</td>
<td>2-10</td>
<td>2-6</td>
<td>2-10</td>
<td>2-6</td>
<td>1-4</td>
<td>2-5</td>
<td>3-6</td>
<td>3-8</td>
<td>2-2 1/2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 6.1**
that the slower students were those who spent less time
studying units any way and that this was the reason that
they were falling behind.

Considering student progress we can examine the record
of time taken between tests and the number of units reached
by each student. This information is displayed in Fig.6.2.
The numbers in the table represent the time taken on each
unit as measured by the number of testing opportunities not
taken up by students, i.e. 0 indicates that a particular
student passed a particular unit on the occasion immediately
following the session when he passed the test for the previous
unit. The unit mean indicates the average time spent by
students between each unit.

From these figures and from the returned feedback sheets
it is possible to determine the units that students took most
time on and also those units that encouraged the greatest
procrastination. It is interesting to note that these are not
the same. Units 5, 12 and 13 were all particularly time
consuming (average time spent >4 hours), but units 3, 5, 6 and
10 caused most procrastination (the average time lag before
students took tests on these units was greater than two weeks).
If the final distribution of the number of units completed by
students is examined, see Fig.6.3, it can be seen that there
were two parts of the course that provided insurmountable
hurdles for some students. These were at units 3 and 11. It
is unexpected that only one unit, unit three, appears to cause
difficulties as assessed by two out of the three different
indices of difficulty, especially when the reported time spent
by students on this unit was one of the lowest. This certainly
suggests that unit three in particular should be investigated
more closely and possibly revised in order to alleviate the
problem.

6.7.2 Long-term formative results

The findings concerning unit three in the last section
<table>
<thead>
<tr>
<th>UNIT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDENT NO.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.38</td>
</tr>
<tr>
<td>03</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>3.08</td>
</tr>
<tr>
<td>04</td>
<td>2</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.50</td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.40</td>
</tr>
<tr>
<td>07</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td></td>
<td>2.60</td>
</tr>
<tr>
<td>09</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td>2.62</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td>2.77</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.09</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2.53</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>3.08</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.80</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.20</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td>2.62</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
<td>2.62</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.63</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.00</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.50</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td>2.92</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.00</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.29</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.80</td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.00</td>
</tr>
</tbody>
</table>

**NOTE**
1. Units are equal to time between each possible testing opportunity, i.e., approximately 1 week.
2. Measured by the time spent between tests.

---

*Fig. 6.2*
### Final Distribution of Students on Units

#### Chemical Physics

<table>
<thead>
<tr>
<th>No. of Units Completed</th>
<th>No. of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1</td>
</tr>
<tr>
<td>3,4</td>
<td>2</td>
</tr>
<tr>
<td>5,6</td>
<td>3</td>
</tr>
<tr>
<td>7,8</td>
<td>1</td>
</tr>
<tr>
<td>9,10</td>
<td>2</td>
</tr>
<tr>
<td>11,12</td>
<td>4</td>
</tr>
<tr>
<td>13,14</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

#### Physics

<table>
<thead>
<tr>
<th>No. of Units Completed</th>
<th>No. of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1</td>
</tr>
<tr>
<td>3,4</td>
<td>2</td>
</tr>
<tr>
<td>5,6</td>
<td>5</td>
</tr>
<tr>
<td>7,8</td>
<td>3</td>
</tr>
<tr>
<td>9,10</td>
<td>2</td>
</tr>
<tr>
<td>11,12</td>
<td>4</td>
</tr>
<tr>
<td>13,14</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

**Fig. 6.3**
could just as easily be included here. They were used partly to modify the present course and partly to influence the course that was to be given a year later.

The questionnaire which was administered shortly before the end of the first term of the course provided the major source of information for this section. The figures ensuing from the structured questions were assembled and the many open-ended comments that students made were grouped together into sections. These formed the basis of a report that was presented to the course organizer shortly before the beginning of the following term. Detailed discussion of some of the questionnaire findings are included in that document (20), but a brief synopsis of them is included here. Emphasis is given to negative feedback, as it is this kind that is most useful in pinpointing changes. A different emphasis would be given in reports to staff, etc.

(i) The overall response rate was 72%. However, if the class is divided in half into 'fast' students and 'slow' students, the response rates become 94% and 54% respectively, i.e. 'slow' students are under-represented in the returns.

(ii) The course feedback sheets whose importance was rated highly by the evaluator received considerable adverse comment. They were rated low (4.1) on the scale of usefulness to the students. Comments included "a nuisance", "no use to students". This result was not unsurprising as most students had effectively stopped using them by this stage in the course and were resistant to being cajoled into returning them.

(iii) The tests were rated at 2.6, however three students rated their usefulness lowly (ratings of 4 or 5). One of these students said that "they are too easy to pass", another thought that the tests had given
him less contact with the staff. The third rejected the entire self-paced system and wanted to return to a conventional lecture course.

(iv) Staff tutors were rated similarly (2.5). A few students thought that the large number of staff in the class would in itself contribute to the success of the course and one pointed to the extravagant use of staff time. A somewhat larger group mentioned that they only consulted with tutors when taking tests and that they did not see them at all on other occasions even when they had problems. Direct observation of the testing sessions suggests that this is so for the majority of students in the class; only about one quarter of the class made more than occasional use of tutors aside from testing.

(v) One area that was particularly studied was that of the amount of time that students were spending on the course and the related issue of the amount of time they thought they were spending on the course. The average time per week that students reported in the questionnaire that they spent on the course was 4.2 hours. This exceeded the time spent as reported on the feedback sheets in the same period by 0.5 hours (not statistically significant). Many comments were made that indicated that students felt that they were spending what they considered to be an excessive amount of time on the course. However, when the time that they were spending was compared to a fair estimate of the time required on the course conventionally, i.e. two hours of lectures, one hour of examples classes, one hour working on problems, then the time spent on the Keller Plan course was not in general excessive. It was clear that students were working more intensively in that time than before and it is possibly for this reason that they overestimated the amount of time that they had spent.
There then arises the question of whether for this reason or others, they spend time on Quantum Mechanics at the expense of other courses. Ten students reported that they did spend more time on other courses. However, nearly all the courses mentioned had a higher nominal loading in the timetable. This could suggest that Quantum Mechanics was receiving more attention from students than its weighting would indicate. It does not indicate that this is due to the method: it was possibly a more difficult course. Three reasons were given equal ratings by the students:

a) the subject is more difficult than others
b) the subject is more interesting
c) the way the course is organized (unspecific)

Most students rejected the following reasons:

d) it is more difficult to learn from books
e) there are too many problems set
f) there is too much information in the course

A small group of slower students (4) did not reject these latter reasons.

70% did report that their time had been more efficiently used in the self-paced method of teaching. Only two thought that their time had been used less efficiently.

In the list of aspects of the course that were rated as useful the most useful parts were: the additional notes provided (1.78), and the textbooks Eisberg (1.70) and Beiser (1.90). The lectures (3.06) and the list of pre-knowledge required (2.78) were rated as least useful. The lectures were interesting, students said, but not useful. The list of pre-knowledge was only found useful when particular difficulties with the unit were found.
When asked their opinion of using a textbook rather than lectures as the basis of the course, students' response was varied. The predominant comment was that it was good but very time consuming. Comments were also made that it placed greater demands on them and that this may be a contributory factor in many students not trying hard on the course. The problem of adapting to this method of working was raised by some students. They said that it was good in principle but that it took time to get used to it. Slower students made less favourable comments on balance. They would have liked to have had lectures in addition to the materials provided. It is this group of students that found most difficulty in adapting to the self-paced style of working.

The main advantages and disadvantages that students reported over the lecture course with tutorials method were as follows (in order of importance):

Advantages related to:

a) self-pacing: 'work when in the mood', 'unable to work at one's own pace', not pushed', 'work at one's own time'.

b) learning: 'more incentive to work', 'work learned more thoroughly', 'each one can concentrate on weak points', 'you are forced to apply yourself', 'learning geared to learning ability'.

c) regular checks on progress: 'makes sure you do some work', 'I understand things as I go along', 'makes one certain of one's progress'.

Disadvantages related to:

a) procrastination: 'tests discourage one from attending', 'easy to slip behind',

b) working from books: 'more difficult to understand', 'less interesting', 'always means that the book is the main source of information'.
c) time taken: 'very time consuming', 'takes more time than usual'.

However, 75% of those responding would like other courses to be run using methods similar to the present one.

These results together with the observational data and the interviews with individual students were used as the basis for proposals for course improvement both for the remainder of the course, and for the following year. Most recommendations could not be implemented in time for the remainder of the course for practical reasons. The substance of these proposals and conclusions were as follows:

"a) The course is too difficult for the weaker students. Only a few students seem capable of completing the course on the recommended schedule.

Proposal: additional notes and some individual tutoring should be considered as immediate remedial action. In the long term the amount of material in the course should be reduced.

b) The course feedback sheets were not useful to the students.

Proposal: if this kind of feedback is desired then either the demand on students should be reduced possibly by taking a sub-set of the group as respondents or the notion of feedback sheets should be incorporated as an integral part of the teaching / learning system the course employs.

c) Lectures were considered to be interesting, but of little use. Weaker students wanted to be lectured to more and many students would like more lectures to be related directly to the content of the course.

Proposal: a few lectures be given concerning the interrelationships of different parts of the course and to give a stimulating overview of the course with the aim of encouraging students to work on the units.
(d) Some weaker students were resistent to presenting themselves for testing. In particular they dislike those tests where they have to expose their ignorance to their peers.

Proposal: steps should be taken to make the testing sessions less threatening and less like examinations. Possibly student tutors could help in this. This problem seems quite fundamental and should be investigated in more detail as it perhaps relates to the overall efficacy of the Keller Plan method of teaching.

(e) The greatest suggestion for change by students related to the times of the testing periods. They did not like them early in the morning and attendance at the 9.00 a.m. period was very low.

Proposal: attempts should be made to schedule the classes at a time that was more attractive to the class.

The attitude scales. Only sixteen students attended the session at the end of the course when the attitude test was administered. These were all students who were fast students, i.e. completed all or nearly all units (11) or who were moderately fast (5). No slow students were represented in the sample. Of the thirteen fastest students in the course only two did not complete the attitude questionnaire. The results are therefore more representative of 'fast' students than even the mid-course questionnaire.

The pre-course and post-course results are plotted in Fig. 6.4. As can be seen very few differences between pre- and post-course ratings are significant. Of the 190 differences there are 17, 6 and 3 significant at the 5%, 1% and 0.1% levels of significance respectively. The number of these that would occur by chance alone are 10, 2, and 0.2 respectively. This indicates that the figures should be examined with caution. Only those differences significant at the 1% level and less will be discussed. These are;
The difference in ratings on the scales shown, before and after the course.

<table>
<thead>
<tr>
<th>Rating</th>
<th>0</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6.4
for the concept QUANTUM MECHANICS there is a change
from vague to precise (.1%), and from unfamiliar to
familiar (.1%); for the concept WAVE-PARTICLE DUALITY
there is a change from unfamiliar to familiar (1%); for
the concept THE UNCERTAINTY PRINCIPLE from unfamiliar to
familiar (.1%); and for PROBABILITY from worthless to
valuable (1%).

The most striking finding is the overall lack of
change. This is partly due to the small sample, in that
it would be expected that there would be a larger number
of significant differences if the sample size was larger.
It does mean that little useful information can positively
be derived from this measure, although it does indicate
that something was being measured reliably.

The end of course questionnaire. Again there was poor
response to this. After two written reminders only nine
students eventually responded. Again they were not
representative of the slow students. All were favourable
towards the course and the method that was adopted. The
reasons given for this were similar to those reported in
the mid-course questionnaire. No consistent reservations
were apparent – different students suggested different
modifications or changes. About one half of the group
felt that there was some small element of competition
involved in the course and a similar number said that the
course had encouraged a limited amount of co-operation both
between students and between staff and students. It was not
possible to draw clear conclusions of any kind from the
responses.

Interviews with tutors. Both staff tutors remarked
that a particular advantage of this method from their point
of view was the close contact with students that they had.
The advantage of having a one to one relationship with a
student was present in the system with, in addition, the
opportunity of the student to choose which tutor he saw.
However, they did feel that it was rather inefficient having two members of staff tutoring at the same time. (This was changed for the later part of the course).

The other positive aspect that they mentioned was that they thought it was of particular benefit to weaker students in that they were obliged to work on the course. One tutor felt that good students would not suffer at all under the scheme, indeed they would probably benefit; the other was uncertain as to its possible effects on these students.

The examination results. Although it was theoretically possible to compare the exam results using the Keller Plan with those obtained the year before it was not possible to do this practically due to the external constraints placed on the form and style of examination. However, in both papers for the different departments the level of difficulty and the general style was fairly consistent from 1971 to 1972. This was confirmed by the external examiner. In the physics group, 75% of the students performed better on quantum mechanics than on the rest of the paper, as against 58% in 1971.

When considering other evidence it is necessary to note that the 1972 physics group had a lower level of attainment in general than the 1971 group. In 1971, 72% passed quantum mechanics (i.e. marks > 55%) compared to 71% in 1972 whilst the corresponding figures for the rest of the paper were 64% in 1971 and 48% in 1972. Of the ten students who failed in quantum mechanics, all but one either failed the examination as a whole or else passed on the borderline. Three other borderline cases did satisfactorily in quantum mechanics.

In the physical science group's examination, where students had a choice of questions it was possible to see the extent to which they were attracted to quantum mechanics
questions relative to the others on the paper. The percentage of quantum mechanics questions out of the total number chosen was 44% in 1971 and 51% in 1972. These may be compared with the probable value of 37.5% given equal choice and the maximum possible one of 60%.

One unexpected finding was that not all students who were behind with their units, did badly. The course finished in March, and all students were then given the remainder of the units, which they had not reached so that they could continue on their own until the examination in June. While, on the whole, the pattern of units reached by March reflected itself in examination performance, there were four students who did very creditably, although they had officially completed zero, four, five and five units out of 15 respectively. This does suggest that some students were able to work effectively on their own using Keller Plan materials but not the recommended procedures.

6.7.3 The start to a summative evaluation

In such a limited study it is not possible to do more than begin to touch on information useful for a summative evaluation, and even more importantly a single course does not provide sufficient examples of the self-paced method from which a conceptual model of these types of courses can be constructed. However, an attempt was made to record information that seemed relevant to this task. It was done by constructing various hypotheses about the effects of self-paced courses, in particular this course, on the students that were experiencing them.

Sources for the hypotheses were many and various. They ranged from comments of students experiencing the course through course tutors, to academic staff who were hostile and also had little or no experience or knowledge of the system. It is necessary for such an uninformed group to be included as it is such people who will be the future consumers
of summative evaluation reports. If their concerns are not aired then it is less likely that they will consider rationally other empirical findings that may be presented. Information was included from the results provided in section 6.6.3 and selected to cover that information that relates to self-paced teaching as a method rather than related to the details of the course in quantum mechanics as such.

The twenty three hypotheses so derived are included in Appendix V and each includes a statement of the rationale underlying each hypothesis*. They are included in this present thesis as a resource for future work that is being undertaken at the time of writing. At the present time it is not possible to report any research or evaluation that has been completed using this framework.

6.8 Outcomes and Conclusion

The question must be asked: to what extent were the findings reported in sections 6.6.1 and 6.6.2 acted upon in the present and subsequent courses? Some of these actions have already been reported in these sections and in some cases these changes have themselves been investigated further. This course was repeated in a modified form in the following year and it was studied (mainly by another investigator) in similar detail. Many of the problems that were encountered were taken into account when planning the evaluation strategy, particularly the poor response rate. It was possible for this subsequent evaluation to study in greater detail those aspects of the course that had been exposed in the present study as problematic. A brief account of this work is given in Willoughby and Boud (143).

*The detailed formulation of the hypotheses and their rationale took place in conjunction with Ms. L. Willoughby
In the case of the Quantum Mechanics course it can be said that the evaluation findings were all considered by the course organizer when planning future courses. That this did happen can be attributed to at least two main factors. Firstly, the formative evaluation activities were relevant to his needs, interests and concerns about the course, and secondly, that a sufficiently firm relationship was established between the evaluator and the course organizer. This mutual confidence in each other's role of provider and consumer of evaluation information was a critical factor. The course organizer could be assured that problems that he considered to be important were being studied and the evaluator could be confident that his findings would at least be considered by the course organizer. These two factors may not appear as of great significance in an objective evaluation strategy, but the dynamics of the real course situation often predominate.

Changes were made in the organization and presentation of the next Quantum Mechanics course, but it is difficult to determine whether they can be attributed to the evaluation or to some other activity such as public discussion, political pressures, etc. It would seem necessary to be able to deduce the effects of a particular activity in order to judge its efficacy, but unfortunately in this situation it is impossible to do this. It is not possible to abstract the evaluation activities from the innovation or the innovative ideas from the results of the evaluation in any useful way. It is possible to ascertain that the initial idea of running a self-paced course was partly a result of information about self-paced courses in the U.S. becoming available and that this was not determined by the present evaluation; but it would be equally true to point out that the problems of earlier courses were partly diagnosed by evaluation activities in these courses. Innovation and evaluation cannot be separated in this sense. Evaluation cannot take place in isolation from the dynamics and fluctuations in real courses and the criteria for good evaluation cannot be divorced from the criteria for good innovation.
At the time of final writing, (July 1974) the Quantum Mechanics course has undergone a series of evaluations and innovations have been introduced both as a result of these and of other considerations about the subject matter. The evaluation reported here continued in subsequent years with a different evaluator. The results presented have been intermediate and tentative rather than conclusive. It is not possible to demonstrate the validity of most of the findings as they resulted in changes in the course which produced an entirely new situation.

In 1973 an account of the first trial of the Keller Plan Quantum Mechanics course which included evaluation data was published (48) and this prompted many inquiries from teachers interested in running their own self-paced courses. Later that year the materials and the methods of the Quantum Mechanics course were transferred to another institution and they were successfully implemented (22). The summative evaluation activities were extended and a preliminary paper was produced in 1974 (23). Both these are indicators that the Quantum Mechanics evaluation was seen to be of sufficient worth for the organizer of the course, and other evaluators to consider that time and effort should be invested to follow the preliminary study. This is one measure of the success of this study that was definitely not present in the biology study and to a relatively small extent present in the physics laboratory study.
PART III

Chapter Seven : Judgments and Conclusions
7. JUDGEMENTS AND CONCLUSION

7.1 Introduction

Earlier chapters discussed some empirical studies in formative evaluation and emphasised some issues related to evaluation. The final chapter aims firstly to review the empirical studies and examine them to judge their worth. This will be done in terms of Stufflebeam's criteria for the evaluation of evaluations. It is intended that this assessment should help to highlight some of the important factors in the evaluations which made them particularly successful or unsuccessful. The style of evaluation in these studies will be discussed and from considerations about the kind of evaluation role that is appropriate in formative course evaluations, a method will be proposed for evaluators working with teachers in the evaluation of university science courses.

It is not intended that this model replace other models but rather that it might be more suitable for some of the particular problems in the particular context in which the evaluations reported in earlier chapters took place. Finally, some brief conclusions will be drawn about the present state of formative evaluation in university courses and some possible directions it might take with special emphasis on the need to validate the supportive evaluation model.

7.2 Criteria for Evaluation of Evaluations

The three case studies will be briefly discussed with respect to each of Stufflebeam's criteria. The chief strengths and weaknesses of the case studies in terms of
7.21 Scientific Criteria

Internal Validity
In apparently objective, numerical scales such as in the aims questionnaire, tests for validity are well established. Such tests have been applied and the results reported. The indications are that the aims questionnaires have high internal validity.

When the Quantum Mechanics measures are discussed then the question of validity cannot be answered numerically. The notion of validity used was one whereby results from each measure were cross-checked with the results from others. Only those findings that received support from more than one measure were reported as valid.

External Validity
External validity is used by Stufflebeam to refer to the generalisability of the information to groups other than the ones it was collected from. In the context of course evaluation the only generalisability that is of concern is that between the group of students studied and other similar groups who may be taking the course either concurrently or in the future.

This is one area where the measures employed may be weak. However, information about validity with respect to other groups has not, in general, been obtained. The biology aims questionnaire raises the suspicion that that version may be unstable with respect to external validity - there are unexplained inconsistencies from one year to the next. For the laboratory aims questionnaire, data obtained in Course A and Course B in other institutions gives considerable support to the stability of that version. The
Quantum Mechanics questionnaires were much more situation specific and so external validity becomes a difficult concept to work with.

Reliability
Both versions of the aims questionnaire have been subject to split-half reliability tests which support the assumption that they are reliable.

Objectivity
With respect to all the information presented to the clients of the evaluations — the teachers, which overlaps almost completely with that presented here, none of the data was doubted after further examination of the evidence presented. This is a rather weak measure of objectivity though, as it demands a rejection of evidence rather than positive acceptance. A more appropriate measure, which was not applied in most cases, would have been to have judgements of objectivity from other evaluators aware of the present evaluator's biases, and from students who had been subjects of the investigation.

Practical and Prudential Criteria

These criteria are concerned with the reception of the evaluation by the clients. They cannot be applied to an evaluation report without knowledge of the situation in which the evaluation takes place. They are, in this sense, not totally public criteria in the way that scientific criteria are.

Relevance
Stufflebeam (130) believes that "the criterion of relevance is very poorly met by almost all extant efforts", and that evaluative information is likely to be obtained
mostly in areas where instruments are available.

Bridgham (26) has also pointed to this phenomena in evaluation studies and uses an analogy of a dart board to illustrate it. If the evaluation instruments are likened to darts, then it is common, he claims, in course evaluation exercises that the dart board— the problem to be studied— is hung where the darts land, rather than the dart board being the target at which the darts are aimed.

In some respects, the studies reported here that have involved the use of the aims questionnaire method have suffered from a failure in relevance. The technique appears to be quite good at producing information of a particular type. However, this type of information was demonstrated to be inadequate in the laboratory case and incomplete in the case of the biology course. The method was relevant to certain ends, but it can be argued that these ends were pursued too far in the laboratory context. In the case of the quantum mechanics course and in the latter parts of the biology course, a much more eclectic approach was adopted which seemed to meet some of the deficiencies of relevance that can be levelled at the laboratory study.

Importance

With respect to the criterion of importance, very similar considerations apply as in the relevance criterion. Judgement was made in the laboratory according to a single evaluative datum. The evaluator was not aware of the undue importance placed in the aims approach at a sufficiently early time so that changes could be made.

Scope

It is necessary to balance scope against relevance: the more relevant a particular instrument becomes to a particular issue, the less relevant it becomes in contrasting
and comparing a range of issues. It was possible to study the well developed method - the aims approach - in great detail and appreciate its limitations. This was not possible in the softer, less refined methods of Quantum Mechanics.

Credibility

In retrospect, the main failings of the laboratory study can be placed in the area of credibility, or rather, its lack. It became obvious at the stage of submitting the final report to physics staff that the evaluator was identified as an outsider whose concerns and interests did not relate to those of the staff to whom the evaluator was nominally reporting.

Conversely, it was possible in Quantum Mechanics for an atmosphere of mutual trust to develop between evaluator and teacher. To some degree this was because the issues of relevance and importance had been addressed at an early stage, but no matter how potent the methodology, little change would have resulted without the strong relationship and understanding of the tasks of both parties.

Timeliness

The more distant a relationship in education becomes, the more a temptation exists to formalise communications and present finished, polished documents. Such documents are rarely timely. Physics laboratory staff were presented with a lengthy, typed report at the end of the study. The Quantum Mechanics' organiser was provided with short, frequent, often verbal reports of progress. With hindsight, Stufflebeam characterises the former stance as self-defeating; it is probably defeating of all useful action.
Pervasiveness
When the evaluator's client is one person then the pervasiveness problem is minimal. In biology there was one client - the course lecturer, in quantum mechanics the client was the organiser, in the laboratory a client was hardly discernable - no wonder that little resulted.

Efficiency
Of all the criteria for effective evaluation that of efficiency is probably the hardest to judge. It cannot be said that if the task had been tackled differently then it would have been more efficient. Stufflebeam regards this as the last criterion to be met; when there are improper applications in the areas of relevance, significance and scope, then efficiency is of minor importance.

What Kind of Evaluation was Appropriate?
Judgements have been made about the separate empirical investigations that have been presented. Little has been said of the characteristics of the style of evaluation that they represent and the worth of that particular style.

Many of the difficulties and limitations of the studies can, in retrospect, be traced to the role of the evaluator. As discussed in 3.4, the evaluator adopted a teacher-centred role which was described from the point of view of the teacher who is responsible for the course. The approach has many dangers and two kinds of teacher-centred approach should be distinguished. The first kind is one in which the evaluator adopts the teacher's perspective and works as if he were the teacher; the second is one where the evaluator tries to adopt the perspective, but recognises he
is not the teacher and should not try to usurp the position of the teacher - the person for whom he is working. In the first case, there is the danger that the teacher may feel threatened in his role and feel as if he is being manipulated covertly. The author believes this to have been the case in the physics laboratory evaluation. There were compounding political factors, but the problem of role was predominant.

A change in style can be distinguished between the laboratory study and the biology evaluation. The evaluator was not a biologist and therefore it was easier for him not to be tempted to take the teacher's role. However, much of the evaluation was governed by problems that were initially introduced by the evaluator: the suggestion of looking at aims and the idea of measuring attitudes. This definition of the evaluation problem by the evaluator also occurred in the laboratory, but there only added to the role problem that already existed. In the biology case, although the evaluator claimed to be taking a teacher-centred approach of the second type outlined here, in practice he was structuring the problem in terms of concepts that he had imported. Whilst this is not necessarily a bad thing to happen, it does characterise the evaluation style.

There was a further change in quantum mechanics. Many of the problems of the evaluator adopting the teacher role had been resolved in earlier years, and during that time a suitably accepting relationship had developed between the evaluator and the teacher. This allowed the evaluator to engage in a teacher-centred evaluation working as an evaluator alongside the teacher, for the most part being directed to problems of concern to the teacher, but also
being aware of unexpected problems that could arise.

These comments have been made on the assumption that the evaluator is a separate person from the teacher and they are not applicable in the situation where a teacher does his own evaluation. In that case, it is entirely appropriate for the teacher to define his own evaluation tasks and to take whatever perspective he finds appealing. If an evaluator wants to adopt this stance then he should find a course of his own to teach.

7.4 Supportive Evaluation

7.41 Background

Pace, in his 1968 review of the state of evaluation concludes by distinguishing two contrasting cases of an evaluation role (97):

"When the evaluator is basically a teacher, reformer or staff officer to the practitioner and the purpose of evaluation is to improve or change a program or practice, then, the process of evaluation is characterised by:

1. A client-centred orientation - in that the clients specify the objectives (usually with help from the evaluators);
2. A co-operative mode of inquiry - in that the clients or practitioners, in addition to the evaluators, plan, conduct, and interpret the inquiry.

The intended result is decision and action.

But when the evaluator is seen as a neutral social scientist and the purpose of evaluation is information and analysis,
Then, the process of evaluation is characterised by:

1. An independent orientation - in that the range of inquiry includes but is not limited by the client's intended objectives;

2. A collaborative mode of inquiry - in that expertise from relevant disciplines is brought to bear on the design, conduct and analysis of the inquiry.

The intended result is the provision of more complex bases for informed judgement."

Pace confirms the transition from a research-centred view of evaluation that predominated prior to 1968 to a decision-centred view that emphasises research and evaluation as distinct activities that both warrant attention in their own right.

The examples of evaluation presented in the earlier chapters can be seen as following the example of Pace's first case, with the reservations about the separation of teacher and evaluator role already discussed. However, in retrospect, when originally conceived the two types of evaluation were not distinguished. The laboratory aims study tends more towards Pace's second case and it was viewed much of the time in a research perspective. By the time that the Quantum Mechanics' study had been initiated a supportive style of evaluation had begun to emerge which closely follows Pace's evaluation for decision and action.

This thesis has presented my empirical work from 1968 to 1972 which has progressed from a conception of the evaluator as a neutral scientist to one of evaluator in a client-centred supportive role. This change has been manifest
in the uses to which the results of the evaluations have been put.

Information from the aims questionnaire study has been used in the design of other courses by O'Connell and Penton, both within the University of Surrey (95) and currently in another institution. In Pace's terms, this could be called "the provision of more complex bases for informed judgement". The Quantum Mechanics study has provided a continuous stream of information which was and is actively being fed back into the course to contribute to its modification: "decision and action". The results have not been used in other contexts, but the range of methods employed have been adopted and adapted by Willoughby and Bridge (23) in their studies of self-paced courses within the Nuffield Higher Education Learning Project (Physics).

Much of the current evaluation literature is concerned with exploring the implications of evaluation designed for decision and action, and in extending Pace's simple dichotomy. There is now a plethora of new types and new names for evaluations under this banner: responsive evaluation (121), discrepancy evaluation (108), transactional evaluation (110), etc. Whilst not wanting to add unnecessarily to the number of labels that have been invented, the following section outlines a style of evaluation that has emerged experientially from the present work.

In review, the empirical studies presented a style of evaluation that was successful in many ways and obtained information that could have been used for course development. However, it was unsuccessful in certain crucial aspects. These were: firstly, it was not in sufficiently close contact
with the perceived problems as seen by the teachers for whom the evaluations were being undertaken. This resulted in procedures being developed that in some cases did not have the complete backing of the teachers, although at no time was there any overt resistance. Secondly, the fact that the problems that were addressed were not those that were normally generated by the clients, did lead to the accumulation of data and analyses that were sometimes not readily understandable and assimilable by them, so that, although potentially useful information was generated, it was not in a form that was easily handled by hard-pressed teachers.

An Alternative Approach

Indications have been given in Chapter 3 and in the previous section of this chapter of a form of evaluation appropriate for 'evaluation for course improvement' in conjunction with an evaluator. The full implications of a supportive evaluation role can only be judged after much experience of it in a variety of contexts. At present, no single case can be cited that provides a clear model for this activity. However, on the basis of some of the shortcomings of the case studies and the experience of the author in working with staff in several institutions some characteristics of a viable role can be outlined. Attention will be focussed on the process of supportive evaluation, rather than on any particular outcomes in the form of instruments, investigation strategies, etc., as these are likely to vary greatly from case to case.

The stages of the process overlap considerably, but the trend is clearer if separate stages of the process are indicated.
Stage 1

In the initial stages of the relationship the evaluator is concerned primarily with establishing a relationship. This involves lengthy discussion with the teacher or teachers in an attempt to appreciate the perspective of the teacher(s) and how he/they view the teaching problems that they are confronted with, the possibilities and constraints as they see them. Emphasis is placed on being supportive and developing a concern for the person and understanding the views that are presented. A minor function at this stage is the legitimising of the evaluator's role. This occurs in two ways. Firstly, through the way he acts and what he says; and secondly, through the evaluator articulating the kind of working relationship that is being sought and the aims that he expects can be obtained. These will be focussed on the process rather than specific outcomes. It must be clear, and seen to be clear, that the evaluator is not present to propagate any particular teaching method or content into the situation, and that his chief function is one of an enabler: a person who enables the people whose prime responsibility it is to improve teaching to operate in an effective way.

Stage 2

A detailed consideration of the issues is conducted. The teacher's proposed strategies are examined and the implications are teased out. Aspects which the teacher sees as problematic are explored in depth. The planning of a strategy for change in the course is initiated. This is done as a joint activity and the roles of both the teacher and the evaluator and their interaction are agreed. Specific tasks by the teacher: curriculum planning, etc; or the evaluator: survey of student needs, etc; may be assigned and a proced-
-ure for sharing information is established. At the end of this stage fairly detailed arrangements for the course should be ready.

Stage 3

Conjoint activities take place. The course is started: the teacher implements his innovations and monitors them from his position; the evaluator supplies additional information through discussion and interview with students and through observation of the teaching situations. He has access to information unavailable to the teacher by virtue of his role.

Stage 4

A review of progress takes place at various times whilst the course is in progress. The evaluator assists the teacher in working through the information that is available from all sources. He gives support to the teacher and helps him find his own direction for modification and change. He respects the teacher's skills and limitations and does not allow himself to be seduced into taking the initiative in proposing ideas that are his own, but outside his client's view of things.

Stage 5

Further changes are implemented with conjoint monitoring and review. The evaluator progressively withdraws from the situation. Stages 3 and 4 are repeated until agreement is reached that the particular course has been developed with the evaluator sufficiently and that the teacher has the capacity to monitor future activities on his own.

Stage 6

Ideally, the stage is reached where the teacher is fully self-monitoring and can rationally evaluate his own teaching activity. The evaluator is available as consultant and counsellor, but no longer acts directly as an agent of the teacher.
There are various problematic elements in the model for supportive evaluation that has been proposed.

The first concerns the initial likelihood of change on the part of the teacher and his course. The scheme proposed is only likely to be effective in a situation in which the teacher is initially committed to change in his teaching and is open, at least intellectually, to the support and involvement of another person.

The second relates to the effective development of such a relationship. The evaluator has it in his power to easily disrupt and destroy the crucial supportive aspects of the situation that is developing. By acting as a teacher himself he is likely to form a threat to the teacher in critical moments. If the evaluator is committed to any specific ideas of change himself, then he will inhibit the evolution of the ideas that the teacher needs to explore. In many ways, the kind of relationship that is proposed is similar in its initial stages to the relationship in student counselling between the professional counsellor and his client, the student (94)*.

The main reasons for proposing supportive evaluation as an alternative in the situations that were experienced in the case studies are that they both encompass those aspects of the studies that were found to be valuable, and meet many of the criticisms that can be levelled at the earlier approach. It retains the empirical, investigative activities of the evaluator and introduces the element of co-operative planning.

* There are also similarities between supportive evaluation and some of the collaborative role models used in organisational development. See, for example, Clark (36), Gouldner (56), and Schein (113).
and empathy with the teacher's perspective. Much work needs to be done to explore the full possibilities and limitations of this supportive evaluation. It is clear that the traditional notion of the evaluator as a neutral, value-free scientist whose function is to provide 'objective', 'valid' information about a course has not proved to be facilitative of decision-making within the realm of university science courses. Such a person can provide information and interpretations for other purposes but these are beyond the scope of the present discussion.

Many questions remain. Is the supportive evaluation role an appropriate one for all course development and evaluation exercises? If it is not, in what circumstances does it become untenable? And what other strategies will then be suitable?

Can the evaluator have multiple-clients? If so, will there be conflicts? Where does the evaluator's responsibilities lie in this situation?

What training is appropriate for such a person? What skills, both technical and interpersonal are required? Can practitioners of the neutral scientist role be retrained?

The limits of the supportive evaluation as outlined here are clear. It is suggested for situations in which a teacher wants to change his course, or is, at least, dissatisfied with what he is doing, and is willing to invest time and energy to do something about it. It is thought appropriate for situations in which there is only a single client, or in which there is an agreed consensus amongst a small group of teachers. It becomes untenable when the evaluator takes upon the role of
teacher for himself and no longer recognises his main function as that of support.

The responsibility of the evaluator is to his client in so far as this produces no conflict with his basic values. He is responsible to the person, not to an organisation or a role. His primary skills need to be interpersonal, although an awareness of the system and context in which the client works is important.

The question of whether traditional social science researchers can be retrained is not so straightforward. There are difficulties in changing one's perspective on problems and there are constraints due to the present reward system for researchers. It may be that it would be easier to use evaluators trained in very different disciplines and who were not affected to the same degree by the traditions and norms of the academic psychologist or sociologist.

Pace (97) points out that "evaluation cannot be described by a single set of rules", "the characteristics of good evaluation differ depending on what is being evaluated, why, and by whom".

If we can in one context develop a viable evaluative role, then we cannot expect it to be transferable without much examination and much change. It appears that a role has been evolved in one setting. Judgement of its worth is mainly dependent on its survival qualities in the environment in which it has grown.
The limitations of the present studies have been discussed and a proposal has been made for a style of evaluation that has greater potential for course improvement than those used previously. At the moment a gap exists between the two: one method has been criticised and another has been put forward to meet these criticisms. What remains is for there to be some empirical test of the proposed supportive evaluation. This could be of two kinds: one, a direct test, the other indirect, but possibly also illuminative. Firstly, studies could be undertaken which follow the prescription outlined in the previous section. The experiences of these from the points of view of all the parties involved: evaluators, teachers, students, administrators could be collated. It would be expected that the teachers involved would express the most favourable response, but it would be hoped that the response of the other groups would not be unfavourable. Secondly, an examination could be made of the work of other people working as evaluators. Currently there is a substantial interest in evaluation and in terms of sponsorship it is one of the growth areas in education. Many new approaches are being proposed and many of these are being tested. A study of some of the new approaches that are close to those of supportive evaluation and those that are distant, but identified to be successful, would give some indications of the potential of the supportive approach. In particular, the notion of responsive evaluation proposed by Stake (120) has many similarities with the supportive style and some elements in illuminative evaluation (100) shed light on practical activities that might be undertaken as one stage. However, one great difficulty for the empirical testing of any of these strategies is that of criteria
for success. None of them depend on the prespecification of the detailed objectives of the evaluation, and so they cannot be judged on that basis. They do all produce results and information which can form the basis for judgement by someone. One of the difficulties of this, though, is that the choice of judge becomes crucial. It will be necessary to be content with a profile of success and limitations of any approach, and not a unitary score. And that is only reasonable. What is wanted is a strategy that is appropriate for a particular situation and it would be expected that different approaches were potent in different areas. Finally, having obtained a range of judges it can be asked, "what criteria should they use?" An answer to this consistent with the evaluation style considered here is that the criteria should not be prespecified. This, however, introduces the difficulty of knowing when one is successful in any particular area. If the same line of argument is followed then it results in the answer, "we will not know until it happens". This leaves us in the position of uncertainty and ambiguity with respect to evaluation. It seems an appropriate place to conclude.


70. Krathwohl D. R., Stating objectives appropriately for program, for curriculum, and for instructional materials development, Journal of Teacher Education 16 (1), March 1965.


127. Steele J. M., Things as they are: an evaluation procedure to assess intents and practice in instruction, unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1969.

128. Stenhouse L., Some limitations of the use of objectives in curriculum research and planning, Paedagogia Europaea, Vol. 6, pp. 73-83.

129. Stones E. and Anderson D., Educational objectives and the teaching of educational psychology, Methuen, 1972.


144. Wittrock M. C., *The evaluation of instruction: cause and effect relations in naturalistic data*, in M. C. Wittrock and D. E. Wiley (eds), *The Evaluation of Instruction*.


ACKNOWLEDGEMENTS

I would like to thank the people who so generously gave of their time and energy in all the courses that were studied.

I especially appreciate the staff and research students of the Institute for Educational Technology. Their stimulus and critical faculties contributed greatly to my development in the 5 years that I was with them. In particular I would like to acknowledge Professor Lewis Elton and Mr. Sid O'Connell for creating and sustaining the supportive environment in which I grew.

My thanks also go to Mary Pennell and Liz Brown who typed the thesis and to Jean Sargent and Tony Tessier who prepared some of the diagrams.
APPENDIX I

Cell Biology and Genetics: Semantic Differential

March, 1972
NAME: 

DEPARTMENT/COURSE: 

64  65  67  68
The purpose of this study is to measure your perception of certain concepts related to your course, by having you judge these concepts against a series of descriptive scales. You are asked to make your judgement on the basis of what these concepts mean to you. **THIS IS NOT A TEST**, as there are no right or wrong answers, and your responses will in no way influence your assessment in this course.

You will find the concepts to be judged in capital letters, for example:

**INDUCTION WEEK**

Below this concept are a series of descriptive scales against which you will judge the concept. An example of a descriptive scale is:

```
chaotic  _:__:__:__:__:__:__ ordered
```

If you feel that the concept is **very closely related** to one end of the scale you should respond:

```
chaotic  X:__:__:__:__:__:__ ordered
```

**OR**

```
chaotic  _:__:__:__:__:__:__:X ordered
```

If you feel that the concept is **quite closely related** to one or other end of the scale (but not extremely so) you should respond:

```
chaotic  _:__:__:__:__:__:__ ordered
```

**OR**

```
chaotic  _:__:__:__:__:__:X:__ ordered
```

cont......
If the concept seems only slightly related to one side as opposed to the other side (but is not really neutral) then you should respond:

chaotic _:_:_\(\times\):_:_:_:_ ordered

OR

chaotic _:_:_:_:_\(\times\):_:_:_:_ ordered

The direction toward which you respond, of course, depends upon which of the two ends of the scale seem most characteristic of the thing you are judging.

If you consider the object to be neutral on the scale, or if both sides of the scale equally associate with the object, or if the scale is completely irrelevant, unrelated to the concept, then you should respond in the middle space:

chaotic _:_:_:_\(\times\):_:_:_:_ ordered

IMPORTANT

1. Be sure to respond to every scale for each concept
2. Do not give more than one response on a given scale

Work as carefully as you can at a fairly rapid pace. Do not puzzle over individual items. Give your first impressions, the immediate "feelings" about the items.

Below is part of a sample page for you to fill in for practice. Do not spend more than a few seconds marking each scale. Your first idea is what is wanted. You can work faster if you do the following:

First, form a picture in your mind of the concept (in this case "University Learning").
Then, read each scale and make your responses very rapidly.
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>valuable</td>
<td>__ : __ :</td>
<td>worthless</td>
<td></td>
</tr>
<tr>
<td>difficult</td>
<td>__ : __ :</td>
<td>easy</td>
<td></td>
</tr>
<tr>
<td>beneficial for society</td>
<td>__ : __ :</td>
<td>harmful for society</td>
<td></td>
</tr>
<tr>
<td>mysterious</td>
<td>__ : __ :</td>
<td>understandable</td>
<td></td>
</tr>
<tr>
<td>dead</td>
<td>__ : __ :</td>
<td>alive</td>
<td></td>
</tr>
</tbody>
</table>
LOOKING BACK ON THE COURSE

precise  vague 1
familiar  unfamiliar 3
simple  complex 5
easy  difficult 7
chaotic  ordered 9
meaningful  meaningless 11
superficial  profound 13
important  unimportant 15
rational  intuitive 17
interesting  uninteresting 19
valuable  worthless 21
unnecessary  necessary 23
dynamic  static 25
satisfactory  unsatisfactory 27
objective  subjective 29
experimental  non-experimental 31
logical  illogical 33
pleasing  annoying 35
successful  unsuccessful 37
positive  negative 39
classical  modern 41
orientated towards concepts  orientated towards facts 43
never intellectually exciting  always intellectually exciting 45
rewarding  unrewarding 47
opportunity for originality  no opportunity for originality 49
never dull  always dull 51
factual  non factual 53
real  imaginary 55
efficient  inefficient 57
useful to me  useless to me 59
the subject: GENETICS

<p>| Precise | Vague      | 1 |
| Familiar | Unfamiliar | 5 |
| Simple   | Complex    | 5 |
| Easy     | Difficult  | 7 |
| Chaotic  | Ordered    | 9 |
| Meaningful | Meaningless | 11 |
| Superficial | Profound  | 13 |
| Important | Unimportant | 15 |
| Rational | Intuitive  | 17 |
| Interesting | Uninteresting | 19 |
| Valuable  | Worthless  | 21 |
| Unnecessary | Necessary  | 23 |
| Dynamic  | Static     | 25 |
| Satisfactory | Unsatisfactory | 27 |
| Objective | Subjective | 29 |
| Experimental | Non-experimental | 31 |
| Logical  | Illogical  | 33 |
| Pleasing | Annoying   | 35 |
| Successful | Unsuccessful | 37 |
| Positive | Negative   | 39 |
| Classical | Modern     | 41 |
| Orientated towards concepts | Orientated towards facts | 43 |
| Never intellectually exciting | Always intellectually exciting | 45 |
| Rewarding | Unrewarding | 47 |
| Opportunity for originality | No opportunity for originality | 49 |
| Never dull | Always dull | 51 |
| Factual | Non-factual | 53 |
| Real    | Imaginary  | 55 |
| Efficient | Inefficient | 57 |
| Useful to me | Useless to me | 59 |</p>
<table>
<thead>
<tr>
<th>Precise</th>
<th>Vague</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiar</td>
<td>Unfamiliar</td>
<td>3</td>
</tr>
<tr>
<td>Simple</td>
<td>Complex</td>
<td>5</td>
</tr>
<tr>
<td>Easy</td>
<td>Difficult</td>
<td>7</td>
</tr>
<tr>
<td>Chaotic</td>
<td>Ordered</td>
<td>9</td>
</tr>
<tr>
<td>Meaningful</td>
<td>Meaningless</td>
<td>11</td>
</tr>
<tr>
<td>Superficial</td>
<td>Profound</td>
<td>13</td>
</tr>
<tr>
<td>Important</td>
<td>Unimportant</td>
<td>15</td>
</tr>
<tr>
<td>Rational</td>
<td>Intuitive</td>
<td>17</td>
</tr>
<tr>
<td>Interesting</td>
<td>Uninteresting</td>
<td>19</td>
</tr>
<tr>
<td>Valuable</td>
<td>Worthless</td>
<td>21</td>
</tr>
<tr>
<td>Unnecessary</td>
<td>Necessary</td>
<td>23</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Static</td>
<td>25</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td>27</td>
</tr>
<tr>
<td>Objective</td>
<td>Subjective</td>
<td>29</td>
</tr>
<tr>
<td>Experimental</td>
<td>Non-experimental</td>
<td>31</td>
</tr>
<tr>
<td>Logical</td>
<td>Illogical</td>
<td>33</td>
</tr>
<tr>
<td>Pleasing</td>
<td>Annoying</td>
<td>35</td>
</tr>
<tr>
<td>Successful</td>
<td>Unsuccessful</td>
<td>37</td>
</tr>
<tr>
<td>Positive</td>
<td>Negative</td>
<td>39</td>
</tr>
<tr>
<td>Classical</td>
<td>Modern</td>
<td>41</td>
</tr>
<tr>
<td>Orientated Towards Concepts</td>
<td>Orientated Towards Facts</td>
<td>43</td>
</tr>
<tr>
<td>Never Intellectually Exciting</td>
<td>Always Intellectually Exciting</td>
<td>45</td>
</tr>
<tr>
<td>Rewarding</td>
<td>Unrewarding</td>
<td>47</td>
</tr>
<tr>
<td>Opportunity For Originality</td>
<td>No Opportunity For Originality</td>
<td>49</td>
</tr>
<tr>
<td>Never Dull</td>
<td>Always Dull</td>
<td>51</td>
</tr>
<tr>
<td>Factual</td>
<td>Non Factual</td>
<td>53</td>
</tr>
<tr>
<td>Real</td>
<td>Imaginary</td>
<td>55</td>
</tr>
<tr>
<td>Efficient</td>
<td>Inefficient</td>
<td>57</td>
</tr>
<tr>
<td>Useful To Me</td>
<td>Useless To Me</td>
<td>59</td>
</tr>
<tr>
<td>Precise</td>
<td>Vague</td>
<td>1</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------</td>
<td>---</td>
</tr>
<tr>
<td>Familiar</td>
<td>Unfamiliar</td>
<td>3</td>
</tr>
<tr>
<td>Simple</td>
<td>Complex</td>
<td>5</td>
</tr>
<tr>
<td>Easy</td>
<td>Difficult</td>
<td>7</td>
</tr>
<tr>
<td>Chaotic</td>
<td>Ordered</td>
<td>9</td>
</tr>
<tr>
<td>Meaningful</td>
<td>Meaningless</td>
<td>11</td>
</tr>
<tr>
<td>Superficial</td>
<td>Profound</td>
<td>13</td>
</tr>
<tr>
<td>Important</td>
<td>Unimportant</td>
<td>15</td>
</tr>
<tr>
<td>Rational</td>
<td>Intuitive</td>
<td>17</td>
</tr>
<tr>
<td>Interesting</td>
<td>Uninteresting</td>
<td>19</td>
</tr>
<tr>
<td>Valuable</td>
<td>Worthless</td>
<td>21</td>
</tr>
<tr>
<td>Unnecessary</td>
<td>Necessary</td>
<td>23</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Static</td>
<td>25</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td>27</td>
</tr>
<tr>
<td>Objective</td>
<td>Subjective</td>
<td>29</td>
</tr>
<tr>
<td>Experimental</td>
<td>Non-experimental</td>
<td>31</td>
</tr>
<tr>
<td>Logical</td>
<td>Illogical</td>
<td>33</td>
</tr>
<tr>
<td>Pleasing</td>
<td>Annoying</td>
<td>35</td>
</tr>
<tr>
<td>Successful</td>
<td>Unsuccessful</td>
<td>37</td>
</tr>
<tr>
<td>Positive</td>
<td>Negative</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>precise</td>
<td>vague</td>
<td>1</td>
</tr>
<tr>
<td>familiar</td>
<td>unfamiliar</td>
<td>3</td>
</tr>
<tr>
<td>simple</td>
<td>complex</td>
<td>5</td>
</tr>
<tr>
<td>easy</td>
<td>difficult</td>
<td>7</td>
</tr>
<tr>
<td>chaotic</td>
<td>ordered</td>
<td>9</td>
</tr>
<tr>
<td>meaningful</td>
<td>meaningless</td>
<td>11</td>
</tr>
<tr>
<td>superficial</td>
<td>profound</td>
<td>13</td>
</tr>
<tr>
<td>important</td>
<td>unimportant</td>
<td>15</td>
</tr>
<tr>
<td>rational</td>
<td>intuitive</td>
<td>17</td>
</tr>
<tr>
<td>interesting</td>
<td>uninteresting</td>
<td>19</td>
</tr>
<tr>
<td>valuable</td>
<td>worthless</td>
<td>21</td>
</tr>
<tr>
<td>unnecessary</td>
<td>necessary</td>
<td>23</td>
</tr>
<tr>
<td>dynamic</td>
<td>static</td>
<td>25</td>
</tr>
<tr>
<td>satisfactory</td>
<td>unsatisfactory</td>
<td>27</td>
</tr>
<tr>
<td>objective</td>
<td>subjective</td>
<td>29</td>
</tr>
<tr>
<td>experimental</td>
<td>non-experimental</td>
<td>31</td>
</tr>
<tr>
<td>logical</td>
<td>illogical</td>
<td>33</td>
</tr>
<tr>
<td>pleasing</td>
<td>annoying</td>
<td>35</td>
</tr>
<tr>
<td>successful</td>
<td>unsuccessful</td>
<td>37</td>
</tr>
<tr>
<td>positive</td>
<td>negative</td>
<td>39</td>
</tr>
<tr>
<td>Term</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>precise</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>familiar</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>simple</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>easy</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>chaotic</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>meaningful</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>superficial</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>important</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>rational</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>interesting</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>valuable</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>unnecessary</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>dynamic</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>satisfactory</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>objective</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>experimental</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>logical</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>pleasing</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>successful</td>
<td>__________</td>
<td>__________</td>
</tr>
<tr>
<td>positive</td>
<td>__________</td>
<td>__________</td>
</tr>
</tbody>
</table>
APPENDIX II

Cell Biology and Genetics: Aims Questionnaire

March, 1972
The purpose of this questionnaire is to find out your opinions of some possible aims that a course in Cell Biology and Genetics might have. It is part of a continuing research project to study possible ways of investigating university courses.

I would like you to give your name and course below. This information will only be seen by myself and the results of this or any other similar tests that you may have done will remain absolutely confidential.

I would also like to thank those of you who have taken part in the project up to now. When any results are published I will put an announcement in Bare Facts so that you can read them in the library.

David Boud.
Institute for Educational Technology
The following is a list of some possible aims for the Cell Biology and Genetics lecture course. Please read each one carefully, and then rate the extent to which you think that the aim is an important one, or otherwise, in A — the present course, and B — your conception of an ideal course in Cell Biology and Genetics.

Use the scale from 1 (not an important aim) to 5 (a very important aim).

<table>
<thead>
<tr>
<th>Aims</th>
<th>Present Course</th>
<th>Your Ideal Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. to expose a student to a minimal body of factual information which is essential to his future work in this or other parts of the course</td>
<td>1 2 3 4 5 cc5</td>
<td>1 2 3 4 5 cc22</td>
</tr>
<tr>
<td>2. to integrate a recapitulation of 'A' level into the new material</td>
<td>1 2 3 4 5 cc6</td>
<td>1 2 3 4 5 cc23</td>
</tr>
<tr>
<td>3. to stimulate a student to read recommended literature</td>
<td>1 2 3 4 5 cc7</td>
<td>1 2 3 4 5 cc24</td>
</tr>
<tr>
<td>4. to inform a student of sufficient concepts and terminology to enable him to gain further knowledge from the recommended literature</td>
<td>1 2 3 4 5 cc8</td>
<td>1 2 3 4 5 cc25</td>
</tr>
<tr>
<td>5. to stimulate reading outside the lecture content</td>
<td>1 2 3 4 5 cc9</td>
<td>1 2 3 4 5 cc26</td>
</tr>
<tr>
<td>6. to encourage a student to accept that statements about physics and chemistry may be taken as a starting point for the purposes of constructing biological hypotheses</td>
<td>1 2 3 4 5 cc10</td>
<td>1 2 3 4 5 cc27</td>
</tr>
<tr>
<td>7. to give examples of hypotheses which may be advanced to relate biological phenomena to accept 'facts' about the properties and behaviour of molecules</td>
<td>1 2 3 4 5 cc11</td>
<td>1 2 3 4 5 cc28</td>
</tr>
<tr>
<td>8. to give examples of observations and experiments which demonstrate how certain experiments constitute evidence for a particular hypothesis</td>
<td>1 2 3 4 5 cc12</td>
<td>1 2 3 4 5 cc29</td>
</tr>
<tr>
<td>9. to show a student that biology is a developing science in which new observations and concepts are being made and older ideas modified</td>
<td>1 2 3 4 5 cc13</td>
<td>1 2 3 4 5 cc30</td>
</tr>
<tr>
<td>10. to enable a student to make criticisms and discern gaps in knowledge or explanations in what he reads</td>
<td>1 2 3 4 5 cc14</td>
<td>1 2 3 4 5 cc31</td>
</tr>
<tr>
<td>11. to enable a student to suggest (albeit wrongly) alternative ideas to those found in the literature he has read</td>
<td>1 2 3 4 5 cc15</td>
<td>1 2 3 4 5 cc32</td>
</tr>
<tr>
<td>12. to make a student aware of the diversity and variability of biological phenomena and of the different types of explanations that may be put forward to interrelate them</td>
<td>1 2 3 4 5 cc16</td>
<td>1 2 3 4 5 cc33</td>
</tr>
<tr>
<td>13. to lead a student to be aware that descriptions of all structure and function are conceptual models, or hypotheses, that have been suggested by observational or experimental results; so that statements of 'fact' need to be qualified accordingly</td>
<td>1 2 3 4 5 cc17</td>
<td>1 2 3 4 5 cc34</td>
</tr>
<tr>
<td>14. to stimulate the student to attempt to understand new discoveries and interpretations as they occur</td>
<td>1 2 3 4 5 cc18</td>
<td>1 2 3 4 5 cc35</td>
</tr>
<tr>
<td>15. to give a student an interest in exploring the subject further</td>
<td>1 2 3 4 5 cc19</td>
<td>1 2 3 4 5 cc36</td>
</tr>
</tbody>
</table>
C. At the end of the present course you will have an examination. Please give your estimate of the probability of each of the following aims being examined at this time. (I realise that you cannot answer with any degree of certainty, but it is only your subjective opinion that is required). Use the scale from 1 (not likely to be examined) to 5 (great likelihood of being examined).

1. to expose a student to a minimal body of factual information which is essential to his future work in this or other parts of the course

2. to integrate a recapitulation of 'A' level into the raw material

3. to stimulate a student to read recommended literature

4. to inform a student of sufficient concepts and terminology to enable him to gain further knowledge from the recommended literature

5. to stimulate reading outside the lecture content

6. to encourage a student to accept that statements about physics and chemistry may be taken as a starting point for the purposes of constructing biological hypotheses

7. to give examples of hypotheses which may be advanced to relate biological phenomena to accepted 'facts' about the properties and behaviour of molecules

8. to give examples of observations and experiments which demonstrate how certain experiments constitute evidence for a particular hypothesis

9. to show a student that biology is a developing science in which new observations and concepts are being made and older ideas modified

10. to enable a student to make criticisms and discern gaps in knowledge or explanations in what he reads

11. to enable a student to suggest (albeit wrongly) alternative ideas to those found in the literature he has read

12. to make a student aware of the diversity and variability of biological phenomena, and of the different types of explanations that may be put forward to interrelate them

13. to lead a student to be aware that descriptions of all structure and function are conceptual models, or hypotheses, that have been suggested by observational or experimental results; so that statements of 'fact' need to be qualified accordingly

14. to stimulate the student to attempt to understand new discoveries and interpretations as they occur

15. to give a student an interest in exploring the subject further
D. Finally, please rate how successful you think that the present course has been in achieving each of the aims for you. Use the scale from 1 (not successful at all) to 5 (very successful).

<table>
<thead>
<tr>
<th>Aim</th>
<th>Scale</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. to expose a student to a minimal body of factual information which is essential to his future work in this or other parts of the course</td>
<td>1 2 3 4 5</td>
<td>cc60</td>
</tr>
<tr>
<td>2. to integrate a recapitulation of 'A' level into the new material</td>
<td>1 2 3 4 5</td>
<td>cc61</td>
</tr>
<tr>
<td>3. to stimulate a student to read recommended literature</td>
<td>1 2 3 4 5</td>
<td>cc62</td>
</tr>
<tr>
<td>4. to inform a student of sufficient concepts and terminology to enable him to gain further knowledge from the recommended literature</td>
<td>1 2 3 4 5</td>
<td>cc63</td>
</tr>
<tr>
<td>5. to stimulate reading outside the lecture content</td>
<td>1 2 3 4 5</td>
<td>cc64</td>
</tr>
<tr>
<td>6. to encourage a student to accept that statements about physics and chemistry may be taken as a starting point for the purposes of constructing biological hypotheses</td>
<td>1 2 3 4 5</td>
<td>cc65</td>
</tr>
<tr>
<td>7. to give examples of hypotheses which may be advanced to relate biological phenomena to accepted 'facts' about the properties and behaviour of molecules</td>
<td>1 2 3 4 5</td>
<td>cc66</td>
</tr>
<tr>
<td>8. to give examples of observations and experiments which demonstrate how certain experiments constitute evidence for a particular hypothesis</td>
<td>1 2 3 4 5</td>
<td>cc67</td>
</tr>
<tr>
<td>9. to show a student that biology is a developing science in which new observations and concepts are being made and older ideas modified</td>
<td>1 2 3 4 5</td>
<td>cc68</td>
</tr>
<tr>
<td>10. to enable a student to make criticisms and discern gaps in knowledge or explanations in what he reads</td>
<td>1 2 3 4 5</td>
<td>cc69</td>
</tr>
<tr>
<td>11. to enable a student to suggest (albeit wrongly) alternative ideas to those found in the literature he has read</td>
<td>1 2 3 4 5</td>
<td>cc70</td>
</tr>
<tr>
<td>12. to make a student aware of the diversity and variability of biological phenomena, and of the different types of explanations that may be put forward to interrelate them</td>
<td>1 2 3 4 5</td>
<td>cc71</td>
</tr>
<tr>
<td>13. to lead a student to be aware that descriptions of all structure and function are conceptual models, or hypotheses, that have been suggested by observational or experimental results; so that statements of 'fact' need to be qualified accordingly</td>
<td>1 2 3 4 5</td>
<td>cc72</td>
</tr>
<tr>
<td>14. to stimulate the student to attempt to understand new discoveries and interpretations as they occur</td>
<td>1 2 3 4 5</td>
<td>cc72</td>
</tr>
<tr>
<td>15. to give a student an interest in exploring the subject further</td>
<td>1 2 3 4 5</td>
<td>cc73</td>
</tr>
</tbody>
</table>
For each of the following topics in the course please indicate your degree of understanding in terms of how confident you would be of answering a question about it in the CABBS examination. Use the scale from 1 (I do not understand it at all - I would not attempt a question on it) to 5 (I understand it very well - I would choose to answer a question on it).

<table>
<thead>
<tr>
<th>Topic</th>
<th>Do not understand - would not attempt question</th>
<th>Understand very well - would choose question on it</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure of protein molecules</td>
<td>1 2 3 4 5</td>
<td>cc 5</td>
</tr>
<tr>
<td>Electron microscopy of cells</td>
<td>1 2 3 4 5</td>
<td>cc 6</td>
</tr>
<tr>
<td>Composition of the cell membrane</td>
<td>1 2 3 4 5</td>
<td>cc 7</td>
</tr>
<tr>
<td>Active transport</td>
<td>1 2 3 4 5</td>
<td>cc 8</td>
</tr>
<tr>
<td>Function of cell organelles</td>
<td>1 2 3 4 5</td>
<td>cc 9</td>
</tr>
<tr>
<td>The endoplasmic reticulum</td>
<td>1 2 3 4 5</td>
<td>cc 10</td>
</tr>
<tr>
<td>Krebs cycle and oxidative phosphorylation</td>
<td>1 2 3 4 5</td>
<td>cc 11</td>
</tr>
<tr>
<td>DNA Structure and replication</td>
<td>1 2 3 4 5</td>
<td>cc 12</td>
</tr>
<tr>
<td>The Meselson-Stahl experiment</td>
<td>1 2 3 4 5</td>
<td>cc 13</td>
</tr>
<tr>
<td>Mitosis and meiosis</td>
<td>1 2 3 4 5</td>
<td>cc 14</td>
</tr>
<tr>
<td>The Mendelian theory of particulate genetics</td>
<td>1 2 3 4 5</td>
<td>cc 15</td>
</tr>
<tr>
<td>Gene-enzyme relationships</td>
<td>1 2 3 4 5</td>
<td>cc 16</td>
</tr>
<tr>
<td>Crossing-over and genetic recombination</td>
<td>1 2 3 4 5</td>
<td>cc 17</td>
</tr>
<tr>
<td>Defining the gene</td>
<td>1 2 3 4 5</td>
<td>cc 18</td>
</tr>
<tr>
<td>Chemical basis of mutation</td>
<td>1 2 3 4 5</td>
<td>cc 19</td>
</tr>
<tr>
<td>The cistron and the cis-trans test</td>
<td>1 2 3 4 5</td>
<td>cc 20</td>
</tr>
<tr>
<td>Genetic code and protein synthesis</td>
<td>1 2 3 4 5</td>
<td>cc 21</td>
</tr>
<tr>
<td>Bacterial conjugation</td>
<td>1 2 3 4 5</td>
<td>cc 22</td>
</tr>
<tr>
<td>Genetics of viruses</td>
<td>1 2 3 4 5</td>
<td>cc 23</td>
</tr>
<tr>
<td>Regulator and operator genes</td>
<td>1 2 3 4 5</td>
<td>cc 24</td>
</tr>
</tbody>
</table>
Please comment on any aspect of the course you wish, thinking particularly of those parts/methods that you feel could be improved.

Please give your reactions to filling in this questionnaire.

Thank you for your co-operation.
APPENDIX III

Quantum Mechanics: Course Feedback Sheet

1971
<table>
<thead>
<tr>
<th>Section no. in text books</th>
<th>Interest</th>
<th>Difficulty</th>
<th>Missing preknowledge</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boring — V.interesting</td>
<td>V.easy — V.difficult</td>
<td>i.e. things that the text assumed that you already knew, but didn't</td>
<td>(on difficult problems etc.)</td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Approximate time taken for this unit: 

....... hours.

General comments on the text: 

General comment on the problems: 

[Signature]
APPENDIX IV

Quantum Mechanics: Questionnaire

1971
Many of the questions are quite specific and are designed to find your opinions on certain aspects of the course. If, however, you feel that you are not able to express yourself fully in any of these questions please answer the question and add your own comments on the page opposite. If some of the questions do not apply to you, for example, if you did not go to the lectures, please put N/A rather than leaving them blank.

Any information that you give on this questionnaire will be treated as confidential. Individual replies will not be disclosed to Professor Elton or the tutors and will not be used in any way in assessing you.

Finally I would like to emphasize that the improvement of this method of teaching depends on your reactions to it and it is therefore important that you complete this questionnaire.

Please return it by or before Monday, 6th December either in the class or send it to me in the I.E.T.

David Boud
Institute for Educational Technology
<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Too easy</td>
</tr>
<tr>
<td>4</td>
<td>Too difficult</td>
</tr>
<tr>
<td>3</td>
<td>Both too easy and too difficult</td>
</tr>
<tr>
<td>2</td>
<td>Neither too easy nor too difficult</td>
</tr>
<tr>
<td>1</td>
<td>Both too easy and too difficult</td>
</tr>
</tbody>
</table>

Please rate how you found the following, by circling the appropriate number and state briefly your reasons for finding each useful or otherwise.
What is your opinion of the idea of basing the course on a textbook rather than on lectures?

Do you think that you learn as much by this method compared to lecture courses?

About

More the same Less

1 2 3

Do you work in class at times other than when you are being tested?

Yes No

Are there any changes in the organization of the course that would help remedy this?

Yes No

Please give your suggestions:

If you have not yet reached Unit 6 please answer the following questions, if not please turn over the page.

Please outline the reasons that you are behind in the course:
What is the average number of hours you spend each week on this course, (including time spent in class)?

[ ] hours

Do you spend a longer time per week on any course (including attendance at lectures, tutorials etc.)? Yes No

If yes, please list which course(s)

If Quantum Mechanics takes up a longer time than some of your other courses, is this because:

(i) the subject is more difficult than others Yes No

(ii) the subject is more interesting Yes No

(iii) there is too much information in the course Yes No

(iv) the way the course is organized Yes No

(v) there are too many problems set Yes No

(vi) it is difficult to learn from books Yes No

(vii) other reasons (please specify) Yes No

On this course how do you allocate your time. Please give approximate percentages for each activity.

Reading %

Doing Problems %

Do you think that the time you have spent on this course has been used efficiently compared with the time you spend on other courses?

Less About More efficient the same efficient

1 2 3
Lectures
Which of the lectures did you attend, please tick

1. Research
2. The unquiet scientist

Would you like to see any changes in the topics covered by the lectures?

Yes No

If yes, what would they be:

.................................................................
.................................................................
.................................................................

Tests
The tests on the units were of three different types. Please give your comments on each, in particular which helped you most in understanding the subject?

Short answer questions (including problems):

.................................................................
.................................................................
.................................................................

Essay questions:

.................................................................
.................................................................
.................................................................

Group discussion questions:

.................................................................
.................................................................

Do you think that the tests have given you closer contact with the staff than the normal "lectures and tutorials"?

More About Less
Contact the same Contact

1 2 3
Units

Did you find the notes provided useful?  

<table>
<thead>
<tr>
<th>Not very useful</th>
<th>Very useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4 3 2 1</td>
</tr>
</tbody>
</table>

Would you like to see more notes provided in the units?  

<table>
<thead>
<tr>
<th>Many more</th>
<th>Much fewer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4 3 2 1</td>
</tr>
</tbody>
</table>

How many of the problems do you generally attempt?  

<table>
<thead>
<tr>
<th>Most</th>
<th>Some</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

What was the number of the last test that you completed?  

What do you think are the main advantages and disadvantages of this method over the traditional "lecture course with tutorials" method?  

Advantages: 1. .................................................................

2. .................................................................

Disadvantages:

1. .................................................................

2. .................................................................

If you were given a completely free hand what changes would you make in the arrangements for the present course?

Would you like to see other courses using a method similar to the present one?  

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Over the page are listed most of the topics you have covered in your course. It would be helpful if you would indicate how well you think you understand them by ringing the appropriate number which represents your degree of understanding now.

1. Do not understand it at all yet, and really would need to start right again at the beginning.

2. Do not understand it very well and really would need to spend a great deal more time on it.

3. Understand it fairly well, but would need to spend quite a lot more time on it.

4. Understand it well, but would need to spend a little more time on it.

5. Understand it very well, and don't need to spend any more time on it.

Please draw a line underneath the point you have reached in the course.
<table>
<thead>
<tr>
<th>Unit</th>
<th>don't understand at all</th>
<th>understand very well</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Photoelectric Effect</td>
<td>1 2 3 4 5 54</td>
</tr>
<tr>
<td></td>
<td>The Quantum Theory of Light</td>
<td>1 2 3 4 5 55</td>
</tr>
<tr>
<td></td>
<td>The Compton Effect</td>
<td>1 2 3 4 5 56</td>
</tr>
<tr>
<td></td>
<td>De Broglie Waves</td>
<td>1 2 3 4 5 57</td>
</tr>
<tr>
<td></td>
<td>Wave-particle duality</td>
<td>1 2 3 4 5 58</td>
</tr>
<tr>
<td>2</td>
<td>De Broglie wave velocity</td>
<td>1 2 3 4 5 59</td>
</tr>
<tr>
<td></td>
<td>Wave and group velocities</td>
<td>1 2 3 4 5 60</td>
</tr>
<tr>
<td></td>
<td>The Uncertainty Principle</td>
<td>1 2 3 4 5 61</td>
</tr>
<tr>
<td></td>
<td>Applications of the Uncertainty Principle</td>
<td>1 2 3 4 5 62</td>
</tr>
<tr>
<td>3</td>
<td>Atomic Spectra</td>
<td>1 2 3 4 5 63</td>
</tr>
<tr>
<td></td>
<td>The Bohr Atom</td>
<td>1 2 3 4 5 64</td>
</tr>
<tr>
<td></td>
<td>Energy levels and Spectra</td>
<td>1 2 3 4 5 65</td>
</tr>
<tr>
<td></td>
<td>The Franck-Hertz Experiment</td>
<td>1 2 3 4 5 66</td>
</tr>
<tr>
<td></td>
<td>The Correspondance Principle</td>
<td>1 2 3 4 5 67</td>
</tr>
<tr>
<td>4</td>
<td>Time Dependent Schrodinger equation</td>
<td>1 2 3 4 5 68</td>
</tr>
<tr>
<td>5</td>
<td>Interpretation of the Wave Function</td>
<td>1 2 3 4 5 69</td>
</tr>
<tr>
<td></td>
<td>Time independent Schrodinger's Equation</td>
<td>1 2 3 4 5 70</td>
</tr>
<tr>
<td></td>
<td>Energy quantization in the Schrodinger Theory</td>
<td>1 2 3 4 5 71</td>
</tr>
<tr>
<td>6</td>
<td>Solution of Schrodinger Equation for the free particle</td>
<td>1 2 3 4 5 72</td>
</tr>
<tr>
<td></td>
<td>Solution of Schrodinger Equation for the step potentials</td>
<td>1 2 3 4 5 73</td>
</tr>
<tr>
<td>7</td>
<td>Solution of Schrodinger Equation for the square well potential</td>
<td>1 2 3 4 5 74</td>
</tr>
<tr>
<td>8</td>
<td>Solution of Schrodinger Equation for the Harmonic oscillator</td>
<td>1 2 3 4 5 75</td>
</tr>
</tbody>
</table>
The following statements have been made by students on the course, please indicate whether you agree or disagree with them:

There should be more opportunities for us to take tests

Agree  Disagree  76

One reason I am falling behind is that I find it difficult to get up for early periods.

Agree  Disagree  77

I don't like this method because it treats us like kids

Agree  Disagree  78

The tutors expect too much of us

Agree  Disagree  79

I couldn't work up much enthusiasm for Quantum Mechanics no matter how well it was presented.

Agree  Disagree  80

Other comments:

Thank you for your co-operation.
APPENDIX V

Hypotheses about the Keller Plan

Method of Learning

1973

The detailed formulation of most of the hypotheses is due to Miss Lynette Willoughby.
Hypotheses about the Keller Plan method of learning

These hypotheses formed the basis of the evaluation of two Keller Plan courses that have been studied at the University of Surrey. The hypotheses were not conceived as a rigid set of rules but as a continuously modifiable framework from which to work, to allow the greatest amount of freedom to adapt the evaluation to the situations that would be encountered.

They represent questions that may be reasonably asked of this particular innovation. As such they are written not as hypotheses that can readily be falsified, but rather as headings to guide attention and further investigation.

It was hoped that, apart from indicating how far this first set of hypotheses could be accepted or rejected, the evaluation would suggest new hypotheses, that could be elaborated in the light of the findings, as the evaluation proceeded. In fact only hypotheses 8 and 11 were added to the original list, and these not so much in the light of the evaluation findings but due to further thought on the matter when drawing up the questionnaires. It is hoped that discussions with people involved in Keller Plan courses at other universities will help to extend this list, and particularly to broaden its scope.

The hypotheses were drawn up on the basis of available papers on the implementation of the Keller Plan, almost all of which are from the United States; from work done in the previous year on two Keller Plan courses at Surrey University; and from possibilities suggested by critics and advocates of Keller Plan courses.

The hypotheses are given here with a short explanation of the reasoning behind each one, and any comments on each hypothesis and its testability that have emerged from the evaluation. The hypotheses deal with three main areas, students' learning in the subject, their attitude towards the course, and the organisation of the course.
Hypotheses

1. The material will be better retained.
   a) In the short term.
   b) In the long term.

   In a traditional lecture course assessed only by a final examination it is possible to get a reasonably good grade just by a short period of concentrated revision, but the material is probably quickly forgotten. It is hoped that the unit perfection requirement will ensure that the material is properly learnt.

   It has been suggested (1), though, that each unit could be rote-learnt immediately prior to each test, and again quickly forgotten. It is hoped, a Keller Plan course would minimise this problem, and if the course maximises interest in the subject material the problem should become insignificant. (6,7)

2. The time spent by the students on the subject is more efficiently used.

   The student, by working towards given objectives on specified sections of the subject, and endeavouring to master them for the next test, will be making full use of his time. Giving a lecture on a given topic at a given time assumes that each student can summon up an interest in the subject at that moment and can concentrate on the subject. As the student can here decide for himself when to study he can choose the time when his concentration and mood are best. The student may waste much time sitting through lectures that he finds boring, for whatever reason; is unable to concentrate on; and that are largely incomprehensible, possibly due to a lack of understanding of preceding stages of the course. The printed units here would enable the student, in this latter case, to go back over any section of the work that he found he did not fully understand. Lack of motivation, particularly when examinations are far away, can prevent the student from concentrating and so understanding and digesting the material.

   This is a difficult hypothesis to study as some students have to put in more work than others to get the same result. It is also difficult for a student to say just how long has been spent on a subject, and different students will work at different levels of concentration.
3. Effective study habits will be developed by the students.

The nature of the course should keep the students working at a reasonably steady pace on this course. This steady working is considered by many people to be effective in learning, and here is taken as such. (2)

Given the pressure from other courses and the lack of experience of many students in disciplining their own study, there is a strong possibility that rather than keeping the students working steadily this type of course will allow the students to put off doing the work almost entirely.

4. The study habits it is hoped this type of course encourages in the students will transfer to other subjects.

This hypothesis obviously depends on the previous one, but even if that one could be accepted this one would be very difficult to get strong evidence for. Without strict controls, which it would be impossible to impose on the situation, it would be impossible to say how much influence one course had had, amongst all the other influencing factors, even if improved study habits could be seen to have been achieved.

5. The students will spend more time on the subject.

As each students' progress is monitored continuously they may spend more time keeping up with the course. Also if the student becomes interested in certain aspects of the course and studies them in further detail, as the course is designed to allow, more time will be spent on this further study. (5,7)

It is difficult here to ascertain how much time is actually spent on a course unless very strict timetables are kept. The students subjective feelings about how much time a given course takes up do not always relate closely to how much time is actually spent: if particularly interested in the subject they might not realise how much time is being spent, or they might think a lot about having to do a subject, and so feel that it is taking a lot of time, when in fact not much work is done. It is also difficult here to divorce the way a course is run from its subject matter, the latter possibly largely determining the time spent on the course.
6. If additional time is spent on this type of course less time will be spent on other courses.

If students spend more time on this type of course they must either spend proportionately less time on other courses or spend more time on studying in total.

Again the difficulty here is to find out exactly how much time is spent on any one course, and for this hypothesis diaries would have to be obtained for the work done on every course.

7. The students will begin to take more personal initiative in their studies on this type of course.

Most students will be used to being told almost exactly what work to do on a course, and they will not be used to taking much personal initiative in their studies. Thus working almost completely on their own may prove difficult. It is hoped though that the design of this type of course will help them to make more personal decisions about their work and to take a more involved attitude towards their work.

Here, again, the students activities on any one course may be limited by the pressures of the other courses that they are taking, so that, say, initiative encouraged by one course may never be transformed into action.

8. This type of course will give the students closer contact with the staff.

In the traditional lecture situation there is usually little contact between the lecturer and the student; even when a discussion occurs after a lecture it only includes a small percentage of the class. And in the usual tutorial there are usually at least a few students to each tutor. In the Keller Plan tutorial though there is a one-to-one relation between the student and the tutor, and each student's problems can be dealt with individually.

9. The unavoidable personal contact involved in this type of course aids in the learning process.

The learning process is a personal experience that can benefit from interpersonal reactions, as a person can convey so much more than a book and can answer questions immediately. A Keller Plan course may be criticized on the grounds that it depends so much on the written
word, rather than on the personal experience of a lecture. But in fact
the unit test situation, with its one-to-one relationship between student
and tutor, can provide much more interpersonal interaction than can
usually be provided in a lecture-tutorial system. The provision of a
variety of tutors, as there would normally be for a sizeable class,
should avoid any personality clashes. (2)

This hypothesis assumes the previous one, and though the personal
contact is unavoidable in physical terms, providing the student takes the
tests or comes to the tutorial room, there is no guarantee that there is
any real intellectual contact.

10. Student tutors evoke greater response from the students.

The level of identification and understanding between peers is that
much greater than between lecturers and students. The students would
therefore feel freer to expose their lack of knowledge and understanding
to a student tutor. It has been found in various courses run with student
tutors that their relative lack of experience and advance knowledge does
does not detract from their usefulness. (2, 5, 7)

However, student tutors have not been used in this country.
Postgraduate tutors have been used though. The hypothesis can therefore
be used for this case.

11. The students will understand the subject matter more thoroughly.

As each unit of work is being mastered and tested before new work
is attempted there should not be any gaps in the students' understanding
that would mar understanding of the following sections. The test at the
end of each unit will make the student study the unit thoroughly, and so
not skip through pieces, which would leave gaps in their understanding.

The students understanding though will only cover the section of the
course that the student has covered, which might not be very much.

12. The students will have more interest in the subject matter.

The self-pacing aspect of the system should allow students to
take more time over the parts of the course that they find most interesting,
and study these sections beyond the limits of the questions set. It
is thought that this will encourage them to take a greater interest in
the subject, rather than feeling that they must keep up with the pace
of the lectures and stifle any inquisitiveness they may have felt towards
the further implications of the material that is being covered. (2)

Here, again, the pressures imposed by other subjects may prevent the students from manifesting any interest, and thus taking full advantage of the self-paced aspect of the course. Also any feelings of competitiveness between the students in this type of course may prevent them from following through any interest in particular aspects for fear of falling behind.

13. The students will not feel nervous or apprehensive at taking the unit tests.

As the number of attempts made at each test does not affect the students' final mark, apart from taking up time that could be spent on further units, any apprehension at taking the tests, for fear of failure, such as in traditional tests, should be removed.

Inexperience of regular grading, though, may cause students to be deterred from presenting themselves for tests. (2,4)

14. The testing at every stage of the course removes the pressures of the final examination.

Even when the Keller Plan course includes a final examination the confidence it is believed the student will have through his mastery of the subject should relieve the usual nervousness about taking the examination.

This hypothesis assumes that the students are confident of their mastery of each stage of the course, which is not necessarily so. Also the students can be aware, throughout the course, that they are just working up to the final examination, which can be an off-putting prospect.

15. The 'immediate-marking' aspect of the course ensures that the student can see what stage his mastery of the subject has reached and this should give him more confidence in the work and sustain his interest in the subject.

In a traditional lecture-examination system the student may work for the whole year without any idea of how well he is really mastering the subject and without getting any positive results or rewards, or having any misunderstandings dealt with. In that situation the students' interest can easily wane, and the stimulus to work is lacking until a short time before the examination.

This hypothesis is again dependant on the student feeling confident of his mastery of the material after passing a test.
16. The course will encourage undue elements of competition between, or within, certain groups of students.

As the slower students see that they are falling behind they might become discouraged, while the faster students might see this as an opportunity to 'get ahead', so that competitive feelings might evolve. This possibility of competitiveness is facilitated by each student's progress being openly declared by the unit test he is taking.

17. The other courses that the students will be taking alongside the Keller Plan course will have an adverse effect on their attitude towards the unit test.

As the students will usually be taking traditional lecture courses alongside the Keller Plan course their apprehension at taking tests and examinations in the former might have a detrimental effect on their attitude towards the unit tests. This may cause the rate of test taking to decrease near examination times for other courses, especially if this does not correspond to the end of the Keller Plan course.

Here it would be very difficult to say just what was causing any attitude or action from the students. And anyway the extra work load of revision near examination time would cause the students to put off work that did not need doing immediately.

18. The lectures could be eliminated from the course without any detrimental effect.

The lectures that are included in the course are purely motivational and do not contain work that is testable, although a certain amount of preknowledge, covered by the units, may be necessary to appreciate the lecture. The lectures are included to provide a little diversity in the course and stimulate interest in the subject matter of the course and attendance at them is purely voluntary.

Data collected from various (American) courses indicates that lectures on non-examinable material tend to be rather poorly attended, possibly due to the many demands on a student's time, and it has been suggested (5) that the lectures could be dispensed with altogether without the course suffering at all. The lectures in the 1971-2 Quantum Mechanics (Keller Plan) course at Surrey, though, achieved
over one third attendance even though the students considered that the lecture material should have been more closely related to the course material.

This hypothesis is obviously related to a certain extent to the type and quality of the lectures given and is therefore situation specific to that extent.

19. The immediate feedback obtained from the first students taking each test enables the units and tests to be improved in time for the slower students in the same class, as well as for following years.

The units may have unclear points in them, or may need more supplementary text, and the tests may contain ambiguous questions. As it will be the stronger students that reach each unit first they will be able to sort out the worst problems and inadequacies in time for the weaker students, who probably would have even greater difficulty with them. This means that the unnecessary difficulties can be ironed out to save the majority of the students pointless, disappointing effort. (7)

Difficulties could arise here when working with a large class, but as was tried in a course at the University of Sussex *, additional, duplicated handouts can quickly supplement the original units. Even, though, when actual changes are not made to the course the tutors can be warned about difficult sections so that they are prepared for the difficulties that might be brought to them.

20. The 'unit-perfection' requirement is an essential aspect of the system.

This requirement ensures that the first material is understood before new material is tackled, so that when the more difficult subjects are encountered difficulties do not arise through lack of the necessary background knowledge. It has been suggested (1) that each unit need not be passed, but only studied and tested, before the next unit is tackled, but this would mean that the total of the material covered by the end of the course would not necessarily be fully understood, and

* [A second year Advanced Quantum Mechanics course at the University of Sussex, run by Dr M.Richards and Dr P.Unsworth.]
and the added stimulus to work would be removed. There are possibly subjects, though, in which vague familiarity with a large part of the course is more use than complete mastery of only a limited range of the course, and in such a case there seems to be a valid reason for not insisting on unit-perfection. (5)

21. The students will not cheat in the tests.

The approach to the unit tests is such that the students are encouraged to see them as a personal test to help them, rather than a hurdle put in their way. The student will feel more confident of his mastery of the subject and so will be less inclined to cheat in the tests, so that the usual extravagant precautions against cheating can be dispensed with. (3,8,9)

It is difficult here to decide exactly what is meant by cheating. It can be regarded simply as co-operative learning. Much stress is given to this problem in courses in the United States, where usually final grades are given on the basis of units completed rather than a final examination. This question may be inherently less important due to the major difference in examination strategy between British and U.S. Keller Plans.

August 1973
References

1. The results of a questionnaire administered at the end of the first term (December) of the Quantum Mechanics course run on the Keller Plan in 1971-2.

2. 'Goodbye, Teacher....' F.Keller
   Journal of Applied Behaviour Analysis No.1 (spring '68) 1, 79-89

3. Physics Teaching by the Keller Plan at M.I.T. B.Green
   Educational Research Center, M.I.T.

4. Self-paced Instruction in Engineering: A Case Study B.Kien
   IEEE Trans. on Education vol. F 14 no.1 Feb.'71

5. Theory of PSI Evaluated for Engineering Education
   L.Hoberrock, B.Koen, C.Roth, G.Wagner
   IEEE Trans. on Education vol. E 15 no.1 Feb.'72

6. Student Performance in a Keller Plan course in Introductory Electricity and Magnetism S.Austin and K.Gilbert
   Cyclotron Lab, Physics Dept, Michigan State Uni.

7. Science Education for the Seventies F.Keller
   American Association for the Advancement of Science, 137th meeting

8. Integration of Three New Teaching Techniques in an Introductory Computer Course W.Dunn
   IEEE Trans. on Education vol. E 15 no.3 Aug '72

9. Individualized Instruction in Introductory Physics Swartz and Zipfel
   American Journal of Physics vol.40, 1430 Oct '72