INDIVIDUAL STUDY COURSES IN UNDERGRADUATE SCIENCE:

DEVELOPMENT AND EVALUATION

by

W. A. ERIDGE

(In partial fulfilment of the requirements of the PhD degree, University of Surrey, July 1977)
## ERRATA

<table>
<thead>
<tr>
<th>Page</th>
<th>Reference</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>ADD FOOTNOTE</td>
<td>&quot;The supporting documents are available from the author via the Institute for Educational Technology at the University of Surrey.&quot;</td>
</tr>
<tr>
<td>21/3</td>
<td>Paragraph 3 line 4</td>
<td>&quot;... it should be noted that widespread use of ...&quot;</td>
</tr>
<tr>
<td>21/34</td>
<td>Paragraph 2 line 5</td>
<td>&quot;... it was a radical innovation in university science teaching with a vocal following of both experimental psychologists, and engineers, chemists, and physicists.&quot;</td>
</tr>
<tr>
<td>21/25</td>
<td>Paragraph 3 line 5</td>
<td>&quot;... popularity polls ...&quot;</td>
</tr>
<tr>
<td>3/24</td>
<td>Paragraph 2 line 7</td>
<td>&quot;... than traditional courses ...&quot;</td>
</tr>
<tr>
<td>5/1</td>
<td>Paragraph 1 line 7</td>
<td>&quot;... courses in existence ...&quot;</td>
</tr>
</tbody>
</table>

### References
In recent years several teaching methods have been developed in higher education which place less emphasis on the lecture than is usual in undergraduate courses, particularly in the sciences. Instead, students spend much of their time studying the subject matter individually using specially designed materials. One method called the Keller Plan or the Personalized System of Instruction has become particularly widespread, but there are many other methods, and many different interpretations. Investigations into such innovative teaching methods in turn call for new approaches to the methodology of evaluation, and these will also be described.

This thesis examines the growth and present state of such courses in Great Britain, and in the process both examines the lessons to be learnt from this, and also describes the evaluation methods used. This is first done by presenting the evaluation of one particular individual study course in case study form, as this illustrates the extent to which evaluation can aid course development. The experience gained from the early stages of the evaluation of this and other courses is then used in two ways: firstly to develop and to try out a resource package for use by teachers who wish to evaluate their own individual study courses, and secondly to emphasize the need for, and suggest new theoretical approaches to small scale formative evaluation. Finally, the results of evaluation going beyond the provision of data are examined, paying particular attention to the interaction between the evaluator, the teacher, the aims for the course, and the course itself.
INDEX

1. Introduction

2. Literature Survey
   2I The Literature on Individual Study Courses
   2II The Literature on Course Evaluation

3. The Development of Individual Study Courses in the United Kingdom

4. Case Study of the Development and Evaluation of an Individual Study Course

5. The Processes of Evaluation

6. The Outcomes of Evaluation

7. Summary and Conclusions

References
Appendices
CHAPTER 1 : INTRODUCTION

1.1 The Aims of the Research
1.2 The Context
1.3 The Subjects Under Study
1.4 Overview of the Thesis
1.1 The Aims of the Research

The research described in this thesis was undertaken in support of an innovation in undergraduate science teaching. It had two main aims. The first was to investigate the growth and development of a range of undergraduate science courses in which lectures had been largely rejected in favour of guided individual study by students. The second aim was to examine means by which such courses could be evaluated either by a teacher, or by an individual not actively involved in the teaching and learning of the course. In practice it also became important to assess the interaction of these two aims, in other words, to examine the influence of course evaluation upon the educational innovation in question.

I have broken down these main aims into certain specific groups which are tackled in the chapters of this thesis. In an order roughly corresponding to the ordering of the thesis these are:-

- to monitor the number, the nature, and the range of disciplines of individual study courses as they developed in the British Isles and Eire, and gather together the experiences of the teachers concerned.
- to examine the background against which these courses were developed, and relate this to the eventual form of the innovation.
- to examine one such course as it developed over a period of three years using methods of educational evaluation appropriate for its improvement.
- to use this course evaluation as a forum for exploring means of evaluation appropriate to individual study courses in general and to consider their relevance to other relatively small-scale innovations in higher education.
- to produce and put to trial a method by which teachers' evaluations of their own individual study courses can be developed and systematised.
- to use the evaluation approaches described to examine certain key questions about individual study courses, and to consider the longer term outcomes of evaluation in general.
1.2 The Context

The research which I report in this thesis took place in the United Kingdom between September 1973 and September 1976. It concerns only science and technology courses, primarily in universities and polytechnics. The fact that I was a post-graduate student with a background in applied science is also important to an understanding of the relationship between myself as evaluator and the teachers I worked with, for their courses were often outside the sphere of my direct subject specialism.

Although the research was based at the University of Surrey, Institute for Educational Technology, I was funded as a member of the Higher Education Learning Project in Physics, co-ordinated from Chelsea College London by Jon Ogborn. This project, like my research, started solely in physics but quickly diversified to cover a range of pure and applied sciences.

This background also has an influence on the nature of the research, for the techniques used were primarily those of educational technology rather than, say, educational psychology or sociology, although I have drawn on such disciplines.

The purposes of the research indicated a teacher-centred and committed approach. And the fact that at the start of the research individual study courses were extremely novel ventures which were something of an unknown quantity meant that the evaluation was primarily aimed at development and improvement rather than judgment of the worth of what was in any case a rapidly changing teaching method. Chapter 2 will show that Scriven (1967) termed the former aim formative evaluation and the latter summative evaluation, and it is useful to distinguish the bulk of the work as being formative evaluation. Another characteristic of the research which also springs from the nature of its aims is that it is predominantly descriptive and qualitative, although by the end of the thesis the needs for research were changing (as shown in chapter 6).
1.3 The Subjects Under Study

Like the aims of the research, the subject is twofold; the development of a teaching innovation, and the examination of means by which it may be evaluated. This dual focus is reflected throughout the thesis.

The literature survey in chapter 2 goes into each subject, but it is useful at this early stage to delineate the boundaries of the research. In particular, for individual study courses it is important to define the range of courses under consideration, for without this almost any university course, or indeed any educative process might be thought of as an individual study course. The focus is upon whole courses in undergraduate science teaching which are primarily theoretical in nature, and in which the bulk of the subject matter is presented to the student individually via written study guides, textbooks, audiovisual aids, or any other means of individual communication. Thus, to a large extent, the methods to be described are alternatives to the all-pervasive lecture method in science education (Bligh (1971)).

Although in theory many alternative methods to the lecture might be devised, in practice a quite restricted range of techniques have become widespread. One method in particular, the Keller Plan or Personalized System of Instruction has predominated both in the United Kingdom and in the United States of America, and this thesis focusses upon this method which is described fully in chapter 2.

The second subject is the formative evaluation of courses. Chapter 2 examines the many different definitions for the term "course evaluation" in the literature. It also attempts to draw from the debate about procedures of evaluation some understanding of the decisions to be exercised by the evaluator over the range of options open to him. At present it is enough to point out that the focus is evaluation, or judging the value of, or monitoring courses. This is different, although complementary to the assessment of students, since this latter measurement clearly has bearing upon the evaluation of courses. There are other boundaries. A consequence of the particular aims of the research (and indeed, the project within which I worked), and my lack of specialist knowledge in many of the
courses under study was the fact that the work is primarily
evaluation of teaching method as opposed to the intrinsic content
of the course such as the interpretation, organization, or emphasis
placed upon the given subject of the course in question. A similar
boundary exists in the other direction, in that the evaluation
concentrated upon innovative courses which are just one element
of the broad university setting. Little attempt was made, nor
was it possible within realistic constraints, to examine the total
context of the university setting nor the broadest impact upon
this of the introduction of individual study courses. Whilst it
is argued that these boundaries were realistic and indeed pro-
ductive, the concluding chapters consider their influence upon the
research.

1.4 Overview of the Thesis

The six remaining chapters are as follows:

Chapter 2 Literature Survey
Chapter 3 The Development of Individual Study Courses
Chapter 4 A Case Study
Chapter 5 Methods of Evaluation
Chapter 6 Outcomes of Evaluation
Chapter 7 Conclusions

This section is intended to describe briefly the content of
each chapter and show how they inter-relate, in order to help the
reader see the pattern of the research.

The dual aims of the thesis run throughout, but the emphasis
varies from chapter to chapter. Chapter 2 treats both equally, in
two parts corresponding to the literature on individual study courses,
and on course evaluation.

Following this, chapters 3 and 4 are primarily about individual
study courses. The development of individual study courses reported
in chapter 3 refers only to those in the British Isles. It starts
with details about the handful of courses which pioneered the approach
in the early 1970's, many of which I studied at first hand.
However, the number grew briskly, and in 1975 a survey was carried out by means of questionnaires and interviews to each of the 47 such courses which were then located. A report of this occupies the majority of chapter 3.

Chapter 4 provides continuity throughout the period of research, and a link with the work on evaluation reported later on. It describes the development and evaluation of one of the earliest individual study courses in the United Kingdom. This was the first year Particle Mechanics course at Surrey University which I evaluated over the three year period of the research.

Chapter 5 considers methods of evaluation. It builds upon earlier chapters which show how experience with techniques of evaluation grew up alongside the growth of individual study courses. Two issues are addressed; firstly how this background can be used to provide a means by which teachers can evaluate their own courses, and secondly what has been learnt about this process of small-scale evaluation by an outsider. In the first, the result is quite practical; a resource package for use by teachers in course evaluation. In the second, the result is a theoretical discussion about how far existing approaches to course evaluation can be drawn together into a new emphasis suited to the course and course components which are the subject of this thesis.

Chapter 6 discusses the outcomes of evaluation. It starts by considering the results of applying the theoretical re-emphasis to evaluation derived in the previous chapter, but continues to examine the outcomes in a wider sense. For example, it draws evidence from chapters 3 and 4 about the influence of formal evaluation upon teachers' goals for an innovatory teaching method, upon students' reactions to teaching, and upon the avenues along which individual study courses have, and have not been developed.

The concluding chapter 7 gathers together many of the generalisations which spring from the research, and considers whether this research represents an appropriate way of "supporting teaching for a change" (Nuffield Foundation (1975)).

Reporting work on course evaluation presents many problems, and two in particular have been faced in this thesis. One is the extreme diversity of the data upon which evaluation findings are
based, since this often occupies many pages of questionnaire responses, field notes, or interview transcripts. The problem is quite simply, proving an assertion or finding, and is particularly acute in the case study in chapter 4. The solution adopted is as follows. Within the text of the chapter appears a list of the evaluation procedures used, and as far as possible parts of the findings are keyed to the instruments upon which they depend. Each of the instruments listed is then either reprinted in the appendix or available as a supporting document. Beyond this the analysis and the raw data are also to be regarded as supporting documents to this thesis. This problem occurs elsewhere, for example, in the case of the various trial drafts of the evaluation resource package.

The second difficulty of writing about evaluation is that theory can easily become divorced from practice and evaluation may seem to be an end in itself, rather than a means by which courses can be developed or judged. In the literature survey (chapter 2), this has been tackled by placing a discussion of the subject of the evaluation before a discussion of the evaluative process.

This accounts for the ordering of chapters of the thesis which is not wholly chronological. However, one chapter which is correctly placed in time sequence is chapter 6, and it is important to bear in mind that this reports work done at the end of the period of research which is in parts tentative and preliminary, often aiming to point out areas into which future research might profitably grow.
CHAPTER 2: LITERATURE SURVEY

2.1 Introduction

PART I: THE LITERATURE ON INDIVIDUAL STUDY COURSES

2.2 The Early Literature on Individual Study Courses
2.3 The Keller Plan in Practice: Individual Study Courses in General
2.4 Individual Study Courses in the United Kingdom
2.5 Significant Evaluations of Individual Study Courses
2.6 Conclusions: The Literature on Individual Study Courses

PART II: THE LITERATURE ON COURSE EVALUATION

2.7 Introduction
2.8 Résumé: The Development of Evaluation as a Field of Study
2.9 The Dimensions of Evaluation: Many Ways of Looking at the Task
2.10 Formative Evaluation in Support of Undergraduate Teaching
2.11 Conclusions
2.1 Introduction

Ideally this survey would examine literature with the same dual focus as the research: both a teaching innovation, and the means by which it is evaluated. In practice such literature is almost non-existent, and the majority of the work to be reported is relevant to either individual study courses or their evaluation. Therefore this chapter is in two parts.

Part I is on individual study courses. It is necessarily somewhat lengthy as descriptions of various methods such as the Keller plan or the audio-tutorial scheme will be necessary for some readers. Also unlike part II on evaluation, the literature on individual study courses has rarely been fully reviewed in the past, despite the vast number of publications on the subject which have appeared in the last decade. Part II charts the history of evaluation, and shows up a wide range of theories about its purpose. Beyond this section 2.9 also analyses in terms of dimensions the decisions to be taken about what an evaluator can actually do.

Part I goes from the general to the particular. It starts from the plans that were made for individual study courses in the 1960's, and moves on to examine their results in practice; first mainly in the United States of America, but later in the United Kingdom. At the end of Part I, section 2.5 reviews in depth the few especially significant evaluations of individual study courses.

In both parts of this chapter (and indeed in the thesis as a whole), the basis of the research is important. For its reliance upon applied educational technology rather than pure educational research, psychology, sociology, or philosophy has certain clear consequences, since literature on the practice of evaluation and individual study courses is more central to the thesis than the theoretical background for these practices. Therefore whilst the Keller Plan, for example, is based upon a substantial and not uncontroversial literature on behavioural theories of learning (Skinner (1953)), and similarly course evaluation upon educational psychology and sociology, the prime emphasis is upon course development and evaluation in practice.
2.2 The Early Literature on Individual Study Courses

The literature on individual study courses is huge and predominantly American. As early as 1972 a bibliography by Hess located 150 papers about such courses, and the number has continued to increase since then. Most papers add little to the fund of knowledge about individual study courses, except to the extent that they replicate the findings of other teachers, with other students in other institutions. However, replication is important when individual evidence is slight. Such papers may only reach a limited audience, although the key papers are very widely read (for example, Green (1971)). Therefore, this chapter needs to be selective in the literature about individual study which it details, and an important function is to locate those reports about individual study which have had, and may continue to have greatest influence and significance.

2.2.1 The Beginnings of Individual Study Courses

The boundaries of the term "individual study courses" as it will be used in this thesis were stated in chapter 1. They are full-time science and technology courses in universities and polytechnics, and are alternatives to the traditional means of conveying the material of the course (the "syllabus", "subject matter" or "curriculum") to the students. The crucial distinction is that they do not convey this information to the students primarily when they are working as a group (i.e. in a lecture, a seminar, or certain kinds of laboratory class), but rather the information is mainly conveyed to each student as he works individually on his own.

Such boundaries cut out types of individual study which are not central to the research such as language laboratories, correspondence courses and the Open University.
Also excluded are individual study courses at school level. These have been in existence for many years, the most notable being the Dalton Plan in the U.S.A. which derived from the Winnetka Plan (Burk (1922)). The plans for both these approaches were remarkably similar to those made forty years later by Keller. However, the practice differed considerably.

The distinction in the means of communicating or conveying material made above is important because it suggests whereabouts the beginnings of individual study courses are to be found. For it prompts the questions as to whether teachers have always had available to them means of teaching each student as an individual in a course of realistic size. The answer is no. Only over the past few decades has the potential for duplicating the written page, and that of tape-recording speech economically and conveniently given the ordinary teacher the power of individual communication previously afforded only to the producer of a book, record, or television programme. Until the products of technology were available, a course teacher wishing to convey his own interpretation of a substantial body of knowledge to students had no real alternative to the group teaching situation in some guise. Obviously this development did not take place overnight; the Roneo system of copying is pre-war, and reel-to-reel tape recorders have been used in schools and universities for many years (Embling (1976)). However, both present considerable practical problems, and Elton (1976) has recently pointed out that duplication systems and quite simple audio (and visual) aids such as the cassette recorder have played an important role in the growth of individual study courses. Indeed he suggests that there is still considerable room for their development.

Naturally, there are exceptions; means of teaching students as individuals which were in use long before the advent of the photocopying machine and the tape recorder. The laboratory is one situation in which this is probably true, although it should be noted that laboratory teaching for undergraduates is less than one century old (Bligh (1976)). However, this thesis is mainly concerned with "theory" courses (as opposed to "practical" courses), and in any case laboratory instruction is often arranged in groups of students rather than as an individual venture (Ogborn (1977)).
Thus the beginnings of individual study courses had to wait until the 1950's or 1960's when it became practicable for the teacher to communicate individually with students, first by means of the written word, but later by means of audio-tape, video-tape, and even computers. However, the limiting effect of extra cost is still present, and it is unlikely that the latter techniques could realistically take on the major part of the teaching in an entire university science or technology course.

A prerequisite for the development of individual study courses which was far more widely noted was not part of the products of technology, but the processes of learning. The main stimulus for the growth of interest in devising individual study courses was the wish to apply the systematic process of teaching indicated by laboratory research on learning. Many of the authors who devised individual study courses in the 1960's took as their starting point the work of learning psychologists (notably Skinner (1953)), including Keller (1968), and Celinski (1968). The second source of the beginnings of individual study is therefore to be found in attempts to develop a "technology of teaching".

As a result the period from 1960 - 1970 saw an unparalleled growth of literature describing new forms of individual study courses in university teaching. The remainder of this subsection charts this growth.

In a series of important review papers, B. Goldschmid and M.L. Goldschmid provide a structure for understanding the growth of individual study courses in higher education (Goldschmid and Goldschmid (1973) and (1976)), although they warn that "to some extent the presentation of these several .... approaches in separate sections seems arbitrary, since there is so much overlap among them: they all emphasise learning rather than teaching, clear goals, active student participation, feedback and evaluation, and individual pacing".

The review starts with programmed learning because "its basic principles have been adopted by the originators of most of the ..... systems described". Briefly, programmed learning
is "based on a carefully sequenced series of statements or 'frames' all designed to crystallise the facts or concepts to be learnt (Brethower et al. (1967)). Test questions are embedded in the programme, and immediate feedback as to the correctness of the student's response is provided".

However, the valuable growth which programmed learning is continuing to show in other settings such as industrial training (Green (1976)) has not taken place in higher education. Critics have emphasised the need to create "optimum conditions for learning without controlling the smallest dimensions of learner practice" (Brethower et al. (1967)). For this reason Goldschmid and Goldschmid state "it would seem preferable not to base an entire university or college course on programmed instruction but rather to use it as one element". This has been the case; programmed instruction has rarely been used successfully as a means of teaching an entire course in higher education. Thus whilst its principles are of key importance to later forms of individual study, its practice falls outside the scope of this thesis. Programmed instruction does however illustrate the need for means of individual teacher-student communication, and in most cases, programmed instruction is presented in printed form, although audio-visual programmes as well as programmes utilising the computer have also been produced, (Hendershot (1967)).

It is almost impossible to put into chronological order the various systems for individual study courses which grew up in the 1960's mainly in North America taking on some or all of the principles of programmed learning. One problem is that even well defined systems with one particular originator are often found to be pre-dated by the efforts of a single teacher working in isolation. (Keller cites several such examples (e.g. the Winnetta Plan, Washburn and Marland (1963)).

The Audio-Tutorial was one of the earliest systems of individual study (Postlethwait et al. (1970)). It was devised and first used in a botany course at Purdue University in the United States of America in 1961. Its originator was a biologist S.N. Postlethwait who described the growth of his scheme
As follows:

"The central point of this approach is the individual audio-tutorial booth (or 'self-instructional learning carrel'), equipped with programmed audio-tapes designed to direct the student in various types of learning activities, visual aids and, perhaps, specimens with which to experiment. Usually the taped presentation is not just a lecture but contains a program of activities sequenced so as to produce learning in the most efficient way. These activities may include lectures, reading and/or examining appropriate materials, doing experiments, and watching movies. The study carrels are generally open most of the day, an instructor is available for consultation, and students are permitted to proceed at their own rate."

Postlethwait's scheme was pragmatic, as it was intended to overcome practical problems associated with teaching biology to large classes. There is only a little later evidence that he was influenced by educational research. And it is perhaps because of this emphasis that the Audio-Tutorial system has become so diverse. (One development which is commonly associated with it, and indeed Postlethwait himself, is the Minicourse (Postlethwait and Russell (1971)). Goldschmid and Goldschmid equate the Minicourse to a very short module in the now quite widespread idea of modular instruction.

Thus the Audio-Tutorial system has been particularly flexible. It has accommodated self-contained modules of instruction, and also the concept of progressive mastery of course content, these variations being pioneered by Postlethwait himself. In the United States of America the range of subjects taught by the Audio-Tutorial scheme is wide, although there is a strong concentration of courses in the biological sciences which were its starting point.

In their review of individualised instruction systems Goldschmid and Goldschmid next group learning centres with other more general information retrieval systems. Again growing up in the United States of America in the 1960's, learning centres provide information for learners in a number of media (Groves (1975)). In ambitious cases this may include film, video-tape, tape-slide,
cassette recordings, books, data files and other written resources, models, displays, laboratory experiments, and computer terminals. However, the learning centre generally does not fall within the restriction of this thesis to whole courses, and indeed, one of the key issues raised by such centres is the difficulty of integrating their use into the remainder of a course (Manwaring (1977)). It is important at this stage simply to note the considerable growth of learning centres on both sides of the Atlantic, since they are of considerable potential as one component in an individual study course.

The computer has been the centre of a great deal of teaching innovation (Hooper (1975)). It has many potential contributions to make to undergraduate teaching; that of computer aided instruction is relevant here, although computer management may contribute to certain aspects of an individual study course in the same way that the availability of resources in a learning centre may contribute.

"A computer (terminal), perhaps supplemented by a tape recorder, earphones, slide and motion picture projector, television screen and keyboard, can be utilised as a teaching machine. Other uses of the computer include storage of instructional content, records of the location and nature of additional references and audio-visual materials, information on classrooms, instructors, and students; immediate assessment of students through analysis of student responses fed to the computer by means of a keyboard, an electronic pen or a punched card; simulations of case materials; and prescription of learning tasks tailored to the needs of the individual student whose performance may again be evaluated and new tasks projects or consultations may be assigned". (Goldschmid and Goldschmid (1973)). However, also like learning centres, it is uncommon for C.A.I. to be the sole means of teaching a whole university course, even in the United States of America. For this reason C.A.I. does not feature strongly in the remit of this thesis, although it will become clear in chapter 3 that C.A.I. can be fitted into individual study courses in a particularly appropriate fashion.

In the latter part of the 1960's three new systems of individual
study courses emerged in quick succession. Each had been in its formative stages for some time so chronological sequencing is based somewhat arbitrarily upon the date of the first defining paper on each. The first was a system called Contingency Management, the second the Keller Plan, and the third Individually Prescribed Instruction. All originated in the United States of America, placing considerable emphasis upon their application of learning theory to the teaching process.

Contingency management is in fact not necessarily a teaching method, nor one that involves individual study at all. Its title is that of a principle of learning, and the procedures which spring from it may comprise any method by which this principle is put into practice. The "Premack Principle" (Premack (1959)) involves "more preferred activities being used to reinforce less preferred ones" (Lloyd et al (1972)), i.e. the former are "contingent" on the latter. Between the mid 1960's and the early 1970's a number of courses, predominantly in psychology grew up under the titles "contingency management" and "contingency contracting". However, as well as the application of a principle that many teachers would regard as obvious, these also tended to be individual study courses. (For example, McMichael and Corey (1969) and Malott and Svinicki (1969).) The detailed methods varied between teachers, so that the approach was quickly embraced by a far more clear cut method: the Keller plan, which was to prove the most significant method by a considerable margin.

In 1965 F.S. Keller and J.G. Sherman started teaching introductory psychology in two different American universities by a method which became known by a multiplicity of different titles including the Keller Plan, the Personalised System of Instruction, and Self-Paced Study. The scheme was originally planned for an entirely new institution of higher education in Brazil (Keller (1974)), but political changes in that country meant that reports of the first full trials of the Keller Plan took place in the United States of America, published in 1965 in the classic paper "Goodbye Teacher ...." (Keller (1968)). The Keller Plan grew rapidly resulting in many hundreds of such courses in the United States of America, and by 1975 it had root
elsewhere in the world including South America and the United Kingdom where at least fifty such courses were in existence then (Bridge (1976)).

Keller himself was a behavioural psychologist and a colleague of B.F. Skinner, and it is clear his primary motivation was in applying this learning theory to the process of teaching. However the basis of the Keller Plan is more diverse than this for the individual study materials involved encapsulate many of the more recent trends in educational technology such as the statement of objectives, the potential for a range of instructional methods, frequent testing and feedback on progress, and heavy reliance upon duplication of the written word. Whilst such components do not contradict Skinnerian learning theory, they do go beyond it (unlike Contingency Management).

However, there soon grew up a group of teachers using the method (mainly physicists and engineers in the United States of America) who accepted the method, and developed it in practice, without necessarily knowing, valuing, or feeling committed to its apparent theoretical bases, for example Pennington (1972). (A detailed description of the Keller Plan is postponed until the next subsection 2.2 2.)

Keller's method of teaching dominated the individual study picture to the extent that it has now been taken to cover both Contingency Management (discussed previously) and the system to be discussed next; Individually Prescribed Instruction. (For example, Corey and McMichael describe their research into "Keller Plan" not two years after their paper which was one of the first on the subject of Contingency Management (1969) (1970)).

Individually Prescribed Instruction was the last individual study system which did not grow up in contact with the increasingly predominant Keller Plan. Like contingency management it is based upon an ideal rather than a method; in this case that the instructional system should be fitted to the individual rather than vice versa. First reports of this scheme were at school level in the United States of America (Glaser (1968)), but it later spread to undergraduate teaching in 1970 (Duda (1970)).
The method is based on four principles:

"Before one can effectively teach a student a concept, it is important to be able to state: (1) exactly what it is you want the student to learn, (2) how you will know when the student has learned it, (3) what the student already knows about the subject to be learned, and (4) what more the student needs to know". (Hosticka (1972)).

Goldschmid and Goldschmid (1973) summarise the procedures involved:

"(1) placement instruments, (2) pretests to be administered before each unit of work, (3) post-tests to determine mastery, and (4) curriculum-embedded tests to measure progress. (In most cases, the scoring of these tests is done by instructor aides.)"

Also like contingency management the fact that I.P.I. requires individual study is somewhat incidental, although in practical terms it would be impossible to use a teaching method which was different for each student that did not use some alternative to group teaching. In higher education I.P.I. has become incorporated into the Keller Plan, gaining features such as self-pacing, progressive monitoring of learning, and even peer tutoring, whilst the Keller Plan has been expanded on both sides of the Atlantic to include some measure of course design to suit student needs (see chapter 3).

2.2 The Keller Plan

The purpose of this section is to describe and analyse the initial definitions of the Keller Plan in the literature, and set the scene for an examination of the practice of Keller Plan in the United States of America, and subsequently in the United Kingdom.

Although his followers wrote a surprising amount about the method, Keller himself wrote surprisingly little. The pattern was as follows. Keller and his co-workers started to formulate the scheme in 1963 when they were presented with the job of setting up an entirely new psychology course at the new University of Brazilia. The result was a system which was the seed of the
Keller Plan, placing the kind of heavy emphasis upon the use of programmed texts and laboratory work that would be expected of a behavioural psychologist in the early 1960's. He described it as follows:

"The education program ...... represents a distillation of many things; ..... the use of programmed instruction where possible; the treatment of textbooks, lectures, conferences, etc. as rewards for passing through various stages of individual study and experimentations, the use of lectures as inspirational rather than truly instructional; the measurement of progress by compilations of things the student has successfully done, rather than by grades on examinations". (Keller (1974)).

This scheme was never put into action in the Brazilian setting it was designed for due to major upheaval in the country. Thus during the early 1960's the plans lay fallow, trials did take place in Brazil but they were troubled by the unrest, and in the United States of America Keller and his American co-worker J.G. Sherman (also a psychologist) used variants of the method on a small scale. Some of these tentative trials, and a report of the Brazilian experience were published between 1963 and 1967 (for example Keller et al (1964)), but they were not particularly significant, except to the extent that they show how the general ideas encompassed in the Brazilian scheme (above) were progressively refined.

By 1965 though, Keller and Sherman had grown sufficiently committed to their refined version of the earlier plans that they based their teaching of a substantial course (for around 100 students) in introductory general psychology upon these refinements at Arizona State University. They also encouraged a small number of other teachers to trial what were to be early prototypes of the Keller Plan.

In 1967 Keller, on the point of retiring from teaching, described the system which he had developed out of the early Brazilian plans, and which he had been using for the past five semesters, at a meeting of the American Psychological Association. This address was subsequently published, and represents the
second significant; indeed the most significant paper about the Keller Plan (Keller (1968)).

Almost ten years later this paper still stands as the starting point of discussion on the Keller Plan. It is in three parts: a description of the plan, an evaluation of its practice, and a comparison with other systems of individual study that were then current.

The method he described was one in which the students work when and where they like, on their own at their own pace. In this context the term self-pacing means that the rate of presentation of course material is controlled by the individual student, although this may be limited by the total time span of the course. (In fact this limit is not as significant in the United States of America, where it is often possible for students to register an incomplete or "I" grade at the end of the allotted time for the course, and complete it the following session.)

In describing his scheme Keller emphasised the importance of "feedback" in the learning situation, and the resulting "reinforcement" that this permitted. It was based, as was the early Brazilian scheme on Skinnerian learning theory, but added to it elements of personal, individual interaction with tutors.

The second key idea in his plan was that students should only pass on to new material when they could show that they have thoroughly achieved the objectives, in Keller's words "mastered" the preceding material. These two elements of progressive mastery and self-pacing are at the basis of the practice of Keller Plan. As Keller (1968) put it:

"This is a course through which you may move from start to finish at your own pace. You will not be held back by other students nor forced to go ahead until you are ready".

"The student must show mastery of each unit (by passing a 'readiness' test or carrying out an experiment) before moving on to the next".

† This description is in parts based loosely upon a summary published in collaboration with D.J. Boud and L. Willoughby (1975). Many such descriptions exist but except for its length, Keller's original has advantages of authority and readability.
Most of the detailed procedures derive from these two basic concepts. Tests must be included to check whether mastery has been achieved. Also, self-pacing means that lectures are impracticable as the main source of information since they impose the pace of the lecturer on the student.

The alternative to lectures adopted by Keller was to divide the course material up into written units which the students work through individually. On completing a unit the student showed his "mastery of each unit (by passing a readiness test or carrying out an experiment) before moving on to the next". (Keller (1968)). To do this the student attends a scheduled test period. During these test periods students take tests, study on their own or with friends, or ask the tutors for tutorial assistance. When the student has completed the test he takes it to a tutor who then marks it with him, discussing it, checking the student's understanding, and assisting in any difficulties. The tutor: "will pass upon your readiness test as satisfactory or unsatisfactory ..... Failure to pass a test on the first try, the second, the third, or even later will not be held against you".

If the student passes the test this is recorded, and the student is provided with the next unit in the sequence to take away and study. Keller did not intend that the considerable amount of testing necessary should be provided by academic staff. Taking his inspiration from military training, his scheme used "proctors":

"A proctor is an undergraduate who has been chosen for his mastery of the course content and orientation, for his maturity of judgment, for his understanding of the special problems that confront you as a beginner, and for his willingness to assist". (Keller (1968)).

In return for this, proctors were paid or received course-work credit. Lectures and demonstrations do have a part to play in Keller's Plan, but they:

"..... have a different relation to the rest of your work than is usually the rule. They will be provided only when you have demonstrated your readiness to appreciate them;
no examination will be based upon them; and you need not attend them if you do not wish. When a certain percentage of the class has reached a certain point in the course, a lecture or demonstration will be available at a stated time, but it will not be compulsory". (Keller (1968)). The second part of "Goodbye Teacher ...." (Keller (1968)) reports the earliest ever evaluation of the Keller Plan. Student reaction was described by citing the response of one "typical" student:

"in comparison with courses taught more conventionally this one demanded a much greater mastery of the work assignments, it required greater memorisation of detail and much greater understanding of basic concepts, it generated a greater feeling of achievement, it gave greater recognition of the student as a person, and it was enjoyed to a much greater extent".

The reaction of one "typical" proctor (undergraduate tutor) was also described, and the picture was equally rosy:

"She referred especially to the satisfaction of having the respect of her proctees, of seeing them do well, and of cementing the material of the course for herself ....".

Next, Keller reported the results of Keller Plan courses in terms of the grades they produce, and learning as measured by examination performance. He describes a final grade profile which shows almost opposite characteristics to the typical normal distribution. Approximately 70% of the students from his course obtain A or B grades; C D and E grades were rare, but withdrawal, incomplete, or fail grades were high (20% of students). However, this does not imply that learning was improved; only that the grading policy was different on Keller's courses, yet Keller does not bring this into his argument. However, a further graphical analysis does attempt to compare learning on a Keller Plan course with that on a traditional course by administering the same mid-term examination to both a Keller Plan class and a class taught the same subject traditionally. Again the evidence is in strong support of the Keller Plan, but the comparison is exceptionally tenuous, as the two courses were in different
institutions, for students of different university levels and abilities.

To conclude on Keller's evaluation of his teaching system it is impossible to judge the validity of the excellent results he claims, and there must be some doubt about the research design and the interpretation.

In the third part of his paper, Keller discusses other individual study schemes which he has experience of. One is significant; that of Ferster and Perrott (1968). They replaced written tests with a short, (typically 10 minute) verbal interview between the tutor (once again, often an undergraduate) and student. In the course of this, the student is expected to describe all that he has read or worked through in the previous unit while the tutor plays the role of "the listener, checker, appraiser, and summariser". (Keller (1968)). Thus use of spoken rather than, or as well as written tests has been taken over by several users of the Keller Plan in subjects where the sole use of a written test was judged to be too limiting.

Up to this point, subsection 2.3.2 has described two of the three key phases of Keller's contribution to individual study courses, the tentative plans for the Brazilian venture, and their refinement into the paper "Goodbye Teacher....." which had such a profound impact. His third, and so far final contribution has been concerned with emphasising the roots of his scheme in behavioural theories of learning, and the danger of modifications which are not based upon these theories. This stance has been most clearly emphasised in the Keller Plan Handbook by Keller and Sherman (1974). In particular he analyses the potentially reinforcing features of the plan, and those which may be aversive. Although not strictly part of a description of the Keller Plan itself, these post hoc analyses by Keller of his system are important here.

Amongst the reinforcers available to students through his plan, Keller pinpoints those relating to (i) study from textbooks; "..... those little explosions of satisfaction or relief from tension that come with understanding ..... Such rewards would seldom be sufficient alone to keep study behaviour
strong .... Textbooks usually lack the built-in reinforcement found in a world of fiction".

(ii) getting ahead unit by unit:

".... each step along the road produces a fraction of the students' final A .... With each successive A, the final A comes closer, and the stronger the students motivation seems to be".

(iii) small work units and self-pacing:

".... lead to a greater density of reinforcement than .... those which typify the lecture system .... they also help the student to use those times for study in which counter attractions do not compete for his behaviour".

(And Keller added to this point:

"Frequent tests, when teacher paced, are probably more aversive to a college student than are his standard mid-term or final examinations - studying for them is almost certain to collide with more rewarding matters".)

(iv) immediate grading of test solution sheets:

"encourages the reinstatement of responses made to questions and to strengthen further those that were correct".

(All in Keller and Sherman (197*).

However, Keller does accept that there are potential aversive features of the Plan. He foresees one aspect which could be threatening for students:

"When a student passes a unit test, for example, his next assignment comes as a reward but also constitutes a threat which can only be removed when the next test is taken. The study questions prepared for each assignment to show the student his objectives also served this function. Each one is an obstacle to be surmounted before it is possible to go ahead. And behind these threats lie bigger ones - for example the threat of lost approval, lost time, or lost assurance of the final A". (Keller and Sherman (1974)).

Thus, in summary, Keller's Plan, and his writings about it fall into three categories: the largely confounded attempts between 1963 and 1965 to devise and run a radically new teaching innovation in Brazil; the period between 1965 and 1968 when this method was
refined, developed and changed; and thirdly, when the method proved its practical worth on a large scale, his attempts to maintain the importance of its basis in learning theory rather than solely pragmatism.

2.3 The Keller Plan in Practice: Individual Study Courses in General

Keller's work prompted a huge literature on the application of the Keller Plan in various settings. This section concerns itself mainly with those papers which break new ground in one of a number of areas.

2.3 1. Pioneering the Keller Plan

In the few years after the publication of Keller's paper there was considerable extension and pioneering to be done; introducing the method into new subjects, new levels of teaching, new institutional settings, and even new countries.

Although Keller had retired from teaching, his principle American co-worker J.G. Sherman continued to teach psychology by the Keller Plan, and report his findings (e.g. Sherman (1971)). However, Sherman's main contribution has not been in educational research into the system, but in consolidating its growth by providing the earliest warnings about the risks of the system (Sherman (1974)), by setting up a newsletter in which experience of the method is shared, and later still, by forming the Centre for Personalized Instruction at Georgetown University (currently the publishers of the "PSI Newsletter").

The method took its first steps amongst teachers of psychology, and then expanded to engineering and then physics. An important pioneering paper was published in 1969 by psychologist J.W. Moore et al., so soon after Keller's founding paper that the title "Keller Plan", nor any other title was available. It is perhaps because of this lack of a "key word" that the paper has gone largely unrecognised, for it describes a well designed trial with results in favour of Keller Plan in biology and philosophy as well
as psychology. (These results are the subject of a later subsection.) Other early courses in psychology, and educational psychology were reported by Ferster and Perrott (1968), and Nelson and Scott (1970).

In 1970 Koen published the first important Keller Plan report by a non-psychologist. His subject was engineering, and he started what was to be an important growth of such courses at the University of Texas in Austin. Subsequently the method was taken up by other engineer colleagues, and ultimately resulted in interdisciplinary growth and interest on the Austin campus which lead to one of the largest Keller Plan research projects yet seen (Stice (1976)). The evaluation of this project is analysed in depth in section 2.5.

However, the next expansion into physics was the most important, for Green's paper in the American Journal of Physics in 1971 became one of the most widely read and quoted papers on science teaching. Again the dates of publication provide a rather artificial history here. In fact the work by Moore et al (1969) had by this time expanded to include a physics course, and trials of the method in physics to be reported later by Swartz and Zipfel (1972), and Pennington (1972) were underway. As in the case of Koen, Green's work was the center of spectacular growth of the method at its particularly prestigious base; Massachusetts Institute of Technology. Five years later this was to come to an equally spectacular end (Hirshi et al (1974)).

In the wake of Green's paper, use of the Keller Plan quickly grew to other subjects, levels, and countries. Chemistry followed physics in 1972 when papers by White et al (1972), then Levis and Wolf (1973), and Micah (1973) appeared. Around the same time Motazed reported a Keller Plan course in mathematics (1971), and Knightly and Sayre described a course in library science (1972). Non-science subjects were the exception, but there were instances for example in foreign language teaching (Stanfield (1974)) and philosophy (Berry (1974)).

Still in the United States of America the level at which Keller's method was introduced ranged from the latter school years to advanced undergraduate and post-graduate courses (Sherman (1974)).
Elsewhere, South America, and particularly Brazil was the first non United States country to develop Keller Plan courses for historical reasons (Bori (1974)). However, the momentum has not been maintained there. Green's work was noticed by European scientists visiting M.I.T. In 1973 Elton et al described the first Keller Plan courses in the United Kingdom (Elton et al (1973)) which were prompted by contacts with Green's work at M.I.T., and the United Kingdom quickly became second only to the Americas in the number of such courses running. (See chapter 3.) The other European country to take up Keller Plan at this early stage was the Netherlands where in 1974 van der Klauw and T.J. Plomp (1974) reported four such courses in introductory mathematics and materials science. In the Netherlands growth of the Keller Plan has been slower than in the United Kingdom, and not without significant problems (see section 2.3 2).

This then was the pioneering that took place while the Keller Plan was still in its early stages. However, from 1973 - 1974 onwards there were some further key steps. Possibly the most important was the establishment and growth of the method in Australia (Avey and Northcott (1975)) where the potential for growth is large. Elsewhere the Keller Plan was also being pioneered in the developing countries such as Nigeria and Botswana (Makhurane (1976)) and also France (Laszio (1976)) and Spain (Casanova (1977)).

2.3 2. Experimenting with the Keller Plan

Many of the teachers who took up the Keller Plan experimented with different features of it, and as a result most courses deviated in some respect from Keller's five key points. Keller was concerned about those who rejected his P.S.I. in favour of what he termed "S.L.I.(something like it)" (Keller (1974)):

"to improve upon this system, or to devise a better plan, an understanding of behaviour theory is important. If it isn't reinforcement theory it will have to be some other".

Courses in the U.S. taught by psychologists and other social scientists tended to follow this lead. However, as the majority
of courses shifted into the physical sciences and technology, the changes and experimentation were increasingly intuitive.

Many of the early experiments were prompted by problems in implementing the system. The first and most vital was what has become known as the "procrastination problem" (Green (1971)). Given freedom of pace, many students made slow progress through the units and had not covered the required number by the end of the course. The contrast between the experiments of psychologists and non-psychologists is particularly sharp here. Significant contributions from psychologists include work by Bijou et al (1976), and Sullivan and Hartley (1971), whilst non-psychologists include Austin and Gilbert (1973), Green (1971), Koen (1970), and Meyers (1970). Both psychologists describe research on the effect of "manipulating incentives" or reinforcers within the course upon student procrastination, whereas the non-psychologists took a wider range of approaches which paid less attention to the reinforcing nature of their incentives, but which in each case reportedly reduced procrastination. Austin and Gilbert first used a course calendar tracing each student's progress openly, and setting a suggested minimum pace. This was to be widely adopted e.g. Widden (1977). Green, perhaps more influenced by psychologists like Keller and Sherman, experimented with the provision of an early final examination as incentive to fast progress. Koen and Meyers adopted particularly pragmatic strategies. Koen used the student's rate of progress through the early units to decide the length of later units, so that a continued pace would result in complete coverage of the course. Meyers effectively eliminated procrastination by making attendance at class sessions compulsory. Neither of the latter approaches have been widely adopted but rewards for steady progress, and a clear record of individual progress have become a part of many Keller Plan courses.

Keller's second key concept, the requirement of progressive mastery has been the subject of little research but considerable debate. However, a number of factors stand out. The first is that required levels of mastery vary considerably from course to course both in the U.S.A. (Stice (1976), and in the U.K. (Bridge (1975)).
An early experience in the U.K. reported by Willoughy and Boud (1973) was that "many (students) do not feel that they have fully mastered the material when they pass the test". The case study in chapter 4 examines this, and suggests that standard-format tests may contribute to the students' feeling of confidence. However, it has been widely recognized that "self-paced courses emphasise a problem found in most university courses, but not usually evident. This is that too much material is contained in each course to enable it to be mastered fully". (Bridge (1975)).

Much of the literature on mastery centres around the meaning of the term itself. Sherman (1976) stated:

"Mastery learning does not mean exploring a subject deeply, or having all the knowledge possessed by the expert".

But this is the interpretation of the term made by certain critics (Young (1973)), and Sherman's definition is not entirely satisfactory:

"It does mean the opposite of sloppy, whatever the level. It involves learning well everthing that is assigned, without implications that the assignment is elementary or advanced or that the course is a survey or more intensive".

The remainder of the literature on experimentation with the Keller Plan is concerned with diversifications. One important aspect has been increasing the scope of the system to include laboratory work and computer aided instruction. As early as 1971 Wagner reported incorporating C.A.I. into self-paced courses, D'Arruda (1972) described a system whereby each unit had both laboratory and theoretical components, and in the U.K. Black (1977) organised a course in which laboratory and theoretical units alternated.

This section has shown that in the space of nine years since Keller's original paper (1968), and despite cautions from the originators, each of Keller's five defining features have been modified. Concluding a review on research in this area in 1975 Boud, Bridge and Willoughby stated:

"The most general trend is that of diversification. This
is demonstrated by the fact that the term PSI no longer refers to a unitary method. It is now a generic term used by teachers, some of whom only accept a little of the plan put forward by Keller, and fit it in with their own ideas and needs, whilst others accept and apply all of the plan".

In the remainder of this thesis the generic term "individual study courses" rather than PSI will be used (the latter is rather uncommon in the U.K.). However, where individual study courses deviate little from Keller's blueprint they will be referred to as Keller Plan courses.

2.3 Lessons about Individual Study Courses in General

At least as important as research into the effectiveness of individual study courses has been the accumulation of literature in three areas on how they may be put into practice, and what pitfalls are to be avoided.

Firstly, the "Instructor Manual for Development of a Personalized Instruction Course" by Born (1970), a paper by Green "Fifteen Reasons not to use the Keller Plan" (Green (1974)), and a lengthy essay by Sherman (1974) provide extensive guidance on the implementation of individual study courses.

Secondly, workshops about the Keller Plan have been held throughout the world, and proponents of the Keller Plan have been keen to use the method to teach their colleagues about its implementation. They have been run mainly from the Centre for Personalized Instruction in the U.S.A., and the Institute for Educational Technology in the U.K. The latter courses and their evaluation have been reported by Stace (1976).

Finally, and perhaps most importantly there are lessons to be drawn from cases in which individual study courses have failed in some way. Three highly significant papers report such failures. Sherman (1974) stated:

"the problems fall into three groups: (1) problems inherent in the system - real or imagined, (2) problems created by modifications of the basic pattern and (3) problems created by the fact that while the world of PSI is totally good, the
Predictably, as an originator of the Keller Plan, he saw the greatest danger from modifications to the plan. In particular he stated that "to change the requirement from mastery to perhaps 70% correct", and "to add new additional contingencies" to overcome procrastination were significant sources of failure. Of the problems inherent in the system Sherman lists heavy staff and student workload and administration costs and time, but his emphasis is that these can be overcome by experienced planning.

Two post-mortems on modified Keller Plan courses present a different view. The similarities between the failure of large courses in physics at M.I.T. ("The Rise and Fall of PSI in Physics at MIT" Friedman et al (1975)), and calculus at Twente University in Holland (van der Meer and Plomp (1976)) are striking, and they do not confirm Sherman's analysis. In both cases initially successful courses grew in size (MIT: 100 to 600 students, Twente: 100 to 245 students), ran briefly, and then were discontinued due to faculty pressure. In both cases the growth was rapid and the preparation time for the course units very short; a situation which a lecture course might reasonably cope with, but which was critical for this type of course. However, this was not the key to these failures, nor was it student dissatisfaction. In both cases the key was requiring uncommitted staff to teach on a course in which their role as a teacher would be changed dramatically. In both cases there was amongst teachers a:

"set of opinions, (that) what is important in (physics) teaching is ineffable: it cannot be compressed into objectives, carried out by means of study guides, or justified on unit exams aimed directly at these objectives". (Friedman et al (1975)).

In both cases, teachers drafted in to teach a Keller Plan course they had not themselves planned, actively disliked their teaching role on it. This added certain communication difficulties to the rapid transition, and the change from volunteer teachers and students to conscripts characterised the failures. Both courses deviated considerably from the Keller Plan, but there is no suggestion that this was the source of failure. Indeed at M.I.T.
increasingly major modifications to the system had the effect of improving the course to the extent that it very nearly gained a reprieve.

2.4 Individual Study Courses in the United Kingdom

2.4 1. The Beginnings

As in the U.S.A., in the U.K. two approaches to individual study, the audio-tutorial scheme and the Keller plan seeded the growth of a wider range of courses. And also as in the U.S.A., the audio-tutorial method started earlier in the biological sciences, yet grew more slowly than the Keller plan which started later but grew quickly in the physical sciences and engineering.

Amongst the earliest was a course by G. Manwaring who first introduced the audio-tutorial scheme in first year zoology at Glasgow University in 1969 (Manwaring (1973)). Her work was quickly followed in 1970 by P. Groves (1973) who used a similar scheme in chemistry, and certain later comments apply to both authors. The Keller plan was first used by L.R.B. Elton and his co-workers in introductory mathematics and second year quantum mechanics at Surrey University in 1971 (Elton et al (1973)).

Manwaring based her interpretation of the audio-tutorial scheme around:

"the Self Teaching Laboratory; a multimedia resource centre used as a library ...... (which) contains about 100 learning resources and a variety of media, and is open to about 800 students ......" (Manwaring (1977)).

However, as was often the case in section 2.2 this innovation, and Grove's work in individual study fell near or beyond the boundary of this thesis as it tended not to constitute a whole course:

"There were fifteen units, each being a tape/slide programme needing perhaps one hour's work ...... The resources are essentially on adjunct to, not a replacement of, the formal lecture courses".

The experiment continued, and the self-teaching laboratory
still functions in Glasgow, although the originator no longer works there. It preceded two lines of development. Firstly the notion of resource centres working in active co-operation with courses to provide initial introduction to a subject, to clarify lectures, or for revision has grown to represent a key part of their function (Groves (1975)). Second, and typical of Postlethwait's emphasis, it preceded a wide variety of attempts at individualising instruction which owed little to any one existing plan or system. (For example, teacher-paced courses described below, and the resource based courses at Dundee (Manwaring (1976)).

The Keller plan was used to teach entire courses from the start, although it was "suitably adapted to English conditions" by Elton et al (1973). The modifications included using post-graduates and staff at first rather than undergraduates as tutors, and basing the assessment almost entirely upon the final examination.

It was at this point in 1973 that I started the research reported in this thesis, and shortly afterwards I assumed joint responsibility for two newsletters: the HELP(P) Newsletter (Ogborn and Black (1973)), and the Self-Paced Study Bulletin (Bridge (1975)), which were to provide a source of background and communication amongst many of the science teachers who subsequently took up individual study courses.

2.4 2. Growth, Diversification, and Experimentation

The initiative by Elton at Surrey University described previously led to similar courses there in chemistry (Poole (1974)), and to an experiment in course transfer which involved the materials and procedures of a quantum mechanics course by Elton being taken over first by another staff member in the University (B. Stace), and then by a physicist in a different university (P. Maas, University of Strathclyde) (Boud et al (1975)). This latter procedure overcame many of the preparation difficulties associated with individual study courses. However, it was not repeated elsewhere, although in concluding their evaluation of the scheme, Boud et al report:

"We have been successful in our attempts to show that Keller Plan materials and programmes can be shared successfully
between departments. From our experience, we have learnt that a necessary condition for successful transfer is the incorporation of some flexibility in selection of the material in the course, and the sequence in which it may be taken".

Thus in the U.K., Keller Plan courses started to grow in two disciplines; physics and chemistry (Freemantle (1976)). Mathematics was represented rather less; the early calculus course at Surrey University was discontinued, although later mathematics courses at Southampton (Cohen and D'Inverno (1977)) and the Polytechnic of the South Bank (Bridge and Laurillard (1975)) were the centre of significant growth of the method in these institutions.

As in the U.S.A. many papers appeared which simply described a modified form of the Keller Plan which was put into practice by the author, and his evaluation of its success. However, they did show that modifications to the original plan were more widespread than in the U.S.A. For example, assessment was frequently by final examination only (e.g. Hollinshead (1977)), and undergraduates were very rarely used as tutors or "proctors". One common feature was that within certain limits, (quite often stringent), students encountered the course material at their own rate, and thus although it is probably wrong to give all the courses in this group the title "Keller Plan", they can be grouped under the heading "student-paced courses".

A small fraction of individual study courses reported were not "student-paced" however. Two courses established quite early on were reported by Cryer (1977) and Davis (1977) which, like lectures, presented all students with new work at a regular weekly rate without specifically testing their grasp of the earlier material. In Cryer's course:

"The students are expected to cover the complete course, it being the depth of their knowledge rather than its extent which is limited by their own individual abilities ..... The students' major studies are from a single textbook which they are obliged to purchase. The course involves one tutorial per week for every six students attending, and at present
includes one 'lecture' per week as well'.

These courses have become known as "teacher paced individual study" courses, and their originators have found that they "have many of the advantages of the Keller Plan, but without some of its dramatic contrasts to a traditional system". (Cryer (1977)). More recently Elton has also implemented a teacher paced course for the physical science students studying the student paced course described in the case study chapter 4. And there are signs of similar courses starting in the U.S.A. (Williams (1976)).

Elsewhere Allen (1976) reported a Keller plan course which did not replace a lecture course, but ran parallel to it "to consolidate the lecture course of a conceptually difficult subject (wave mechanics) for .... the weaker section of the student body .... as the departmental ordinary degree course had been phased out". As well as testing this novel approach, Allen also reports one of the few U.K. analyses of examination performance, in this case of students who had (voluntarily) opted for the additional Keller plan course.

"The mean mark of the Keller candidates exceeded that of the others on each section, thus supporting the opinion ..... that the self-selecting group were as a whole more able at examinations than their colleagues. However, while this excess was 18%, 24% and 32% for other courses, it was 94% for wave mechanics. If this behaviour is repeated in their final examinations in the third year, and with future second year classes, ..... the cognitive improvements ..... will be without doubt".

A similar course was run in chemistry at Surrey University by Poole (1974); again the Keller plan was intended to support the lecture course rather than replace it, and again a thorough analysis of the examination results was described. The latter replicated Allen's results, and the course has continued (and indeed expanded) for this reason. It is perhaps predictable that courses of this kind should be among the few to fully report examination result analysis since they demand extra resources which must depend upon improved tangible results.

So far individual study courses in physics, mathematics,
and chemistry have been described. In 1977 Widden described a long established course in engineering science which has led to other courses on similar subject matter (for example, at Southampton University). Widden's course was the nearest to Keller's prototype yet seen in the U.K., being developed in comparative isolation, and it was the first to make full use of undergraduate "proctors". No psychologists in the U.K. have taken up Keller plan teaching, although a psychologist in Dublin has experimented extensively with the method (Baker (1974)). And most recently, and promisingly, teachers of medicine have also started to experiment with the method (Stoward (1976)).

2.4 Future Developments

1977 should produce certain significant developments in the literature on individual study courses, and this section briefly extrapolates from up-to-date U.K. literature to gauge the future of individual study courses.

Up to now books on individual study courses such as the Keller plan or the audio-tutorial scheme have been primarily American (e.g. Keller and Sherman (1974)). A book combining the experiences of many of the originators of these courses in the U.K. was published in May 1977 (Bridge and Elton (1977)), and this may mark a new phase in the history of such courses in the U.K. For until this point the movement has resulted from a chain of personal contacts, and mutual support and advice. However, recent literature makes clear that both the methods included under a generic heading such as individual study courses (or even Keller Plan courses), and the contexts of these courses, are growing too diverse to be centrally co-ordinated. As well as growth into new disciplines, individual study courses increasingly appear in fields of higher education other than the universities, such as polytechnics (see chapter 3).

The signs are that individual study courses will continue to diversify. The Keller Plan may not continue to expand, as experience of the courses Keller termed SLI (something like it) grows to the extent that the plan does not represent the only relatively safe and proven means of teaching students as individuals.
It could also be limited by cost, for without undergraduate proctors, large, and even medium size Keller Plan courses (say 30-50 students) may require more tutor time than traditional courses (Black and Boud (1977)). And it is unlikely that a small group of innovative teachers can bring about the change of attitude amongst university teachers in the U.K. required if proctors are to be widely accepted.

Future literature may show a slowdown of the plans and systems detailed in this literature survey, but they will have acted as the forerunner to a wider range of alternatives to the lecture. They will have shown that "there is more than one way" (Stice (1976)), and provided some evidence for the discussion in section 2.5 about the really essential features of individual study courses.

2.5 Significant Evaluations of Individual Study Courses

There are a small number of especially significant evaluations of individual study courses in the literature. Some are well designed and extensive evaluations carried out by independent researchers, and some are analyses of the data from a wide range of evaluations by teachers reported in the literature. In terms of the number of published papers J.A. and C.L.C. Kulik have been the most prolific in this area:

"The Keller Plan in Science Teaching" published in February 1974 with K. Carmichael
"Effectiveness of the Personalized System of Instruction" published in December 1975, and

These are primarily compilations of published teacher's evaluations, although they do contain original research by the authors.

I was a member of the second group, which was the only attempt at combining first-hand and reported evaluations from the U.K. as
well as the U.S.A.:


Finally, and most recently were two contrasting yet significant evaluations. The first was:


This reported first-hand, the evaluation by independent researchers of 17 different individual study courses at the University of Texas which stemmed from Koen's work reported earlier. The other was:


There have been other summaries of evaluation research and other independent evaluations (e.g. Stoward (1976)), but these encompass almost all of the significant areas of debate. Despite the titles, these evaluations do cover the whole field of individual study courses, for although the authors have tended to start their investigations into the Keller plan, later on the diversity of practice has been acknowledged. Three main areas of evaluation are tackled in the following subsections.

2.5 1. Student Reaction to Individual Study Courses

Almost every published report on individual study courses includes a formal or informal evaluation of student reaction. The U.S. research by Kulik et al; Stice, and Traveggia are all in agreement that this is strongly favourable. The first paper by Kulik et al produced the following findings which were not subsequently challenged:

"The most convincing of all the student rating studies are those that compare a PSI course with a conventionally taught course". (Five studies are then cited which make a methodologically sound comparison. All are favourable. No unfavourable comparisons were located) ....."The attractiveness of the plan to students therefore seems no longer a matter of controversy".

The U.K. evaluations, although generally positive, do
contain a small number of unfavourable overall student reactions (e.g. Freemantle (1976)) as do some U.S. reports (e.g. Hirschi et al (1975)). Boud, Bridge and Willoughby summarise:

"There is much evidence confirming the value of self-pacing. Its main disadvantage is to those students who find it easy to procrastinate ...... There are fewer positive findings about the other basic principle, mastery. It is clear that some teachers require much greater mastery of the units than others, and that students do not always feel very confident of their knowledge of the material, even when they have passed a test on it".

2.5 2. Student Learning on Individual Study Courses

Student learning, in terms of end-of-course performance on some kind of examination is seen as the acid test of many teaching methods. Kulik et al and Traveggia's evaluation surveys assert with increasing confidence that individual study courses with certain characteristics significantly outperform the traditional lecture-tutorial system on examination-tested learning. The independent evaluations reported by Stice differ, and:

"..... are not as favourable to PSI ..... Twelve comparisons were made between PSI courses and their associated 'control' courses ..... In six comparisons the PSI sections did significantly better. There was no difference in five comparisons, and the control class had better achievement in one instance".

Without the resources nor the psychometric background of many of the U.S. psychologists using individual study courses, the U.K. evaluations cannot add significant control studies of learning (see Boud, Bridge and Willoughby, page 25). However, the data which does exist supports Stice's picture of variation rather than Kulik et al and Traveggia's picture of uniform improvement. This subsection examines the disparity of emphasis in this key area.

Results of prior research on comparative learning achievement are conclusive, although perhaps surprising. In 1972 Travaggia and Hedley reported on a comprehensive study of over
40 years of comparative research in teaching methods, including 170 empirical studies of lecturing, group discussions, one-way television instruction, and unsupervised independent study. The result confirmed their earlier conclusion that:

"there is no measurable difference between distinctive methods and media of college teaching when evaluated by students' performances on course content examinations". (Dubin and Traveggia (1968)).

Kulik et al and Travaglia's work indicated to their authors that this impasse had been broken:

"Until PSI came along, all teaching methods seemed to be about equal in effectiveness".

How did they arrive at this conclusion? In the three papers by Kulik et al, and that by Travaglia, the basis was published evaluations by teachers which compared individual study courses with traditional courses. The criteria for selection were stringent, including comparability of experimental and control groups, totally uniform testing procedures between groups, no differential subject loss from the comparison groups, and no differential exposure during the course to test items included in the final exam.

Kulik et al located just five reports in 1974 which met these criteria, but a total of 39 in 1975:

"we found 39 studies which ..... met minimal requirements of experimental design. In 38 of these studies exam performance was better in the PSI course, and in 34 studies the performance difference ..... was great enough to be considered statistically reliable. In one case, lecture performance was slightly better, but the difference did not reach, or even approach, statistical significance".

In 1976 Travaglia located only 14 comparative reports which met his minimal requirements of experimental design. He sorted these according to the extent that it was possible to independently interpret the data (rather than accept the author's analysis). The reports made a total of 28 comparisons. All were favourable towards individual study courses. Six comparisons could be independently interpreted, and in these the difference was
significant at p<0.05.

Traveggia and Kulik et al agreed on their conclusions: "When evaluated by average student performance on course content examinations, the Personalized System of Instruction has proven superior to the conventional systems with which it has been compared".

What is the reason for the difference between the findings of Stice's relatively independent evaluations of student learning, which show a picture of variation of learning gains, and those of Kulik et al and Traveggia obtained by surveying published course evaluations by teachers which indicate that Keller Plan is uniformly superior? Only the third survey by Kulik et al discusses the possibility that their analysis could be flawed. The three flaws they consider possible are that the apparently uniform superiority of individual study courses is due to: "the Hawthorne effect ....." "teaching the test" ..... or "selective publication". They discount the first by observing that long established individual study courses have been reported to produce gradually improving examination results as, in time, they lose their novelty.

Secondly, the possibility that individual study courses appear superior because they train students in test passing (i.e. they 'teach the test') is discounted as this superiority is maintained in retention studies long after the end of the individual study courses, and in studies using quite different test formats.

Thirdly, they argue that selective publication does not cause the weight of significant evaluations in favour of individual study. They accept that the lack of equivocal or non-significant evaluations may be due to the fact that "few investigators submit articles with non-significant results for publication, and fewer editors choose to print such reports". However, they assert that this characteristic of selective publication does not account for the lack of significant comparisons "against" individual study:

"Whenever teaching methods have been compared in the past, the number of significant results favouring one method has been exactly balanced by the number of significant reports favouring the other method. If PSI were just another innovation, we should expect to find many published reports
in which PSI was definitely inferior to the lecture method. Instead no one has reported a result of this kind". (Kulik et al (1976)).

It can be shown that these defences of the analysis by Kulik et al and Traveggia are defective in two respects: (1) the last argument, on selective publication, is inadequate; and (2) the first, on the Hawthorne effect, although sound, ignores a broader issue which was not considered.

Firstly, Kulik et al's analysis of selective publication ignores the intense enthusiasm and commitment on the part of teachers whose evaluations of the Keller Plan (in particular) he bases his survey of learning upon. In the early 1970's the Keller Plan was not "just another innovation". It was a radical innovation in university science teaching taken up by a vocal following, some experimental psychologists, but many others were primarily engineers, chemists or physicists. Comparisons between the Keller Plan and lecture courses involve far more differential commitment than the comparisons "in the past" which Kulik et al refer to, such as those between group teaching and lectures, or totally independent study and lectures. Furthermore these earlier studies were almost all carried out by social scientists conducting educational experiments in which no result of significance would be counted a "failure". (Contrast this with the "failure" of Keller Plan at MIT reported in the previous section.) This differential commitment on the part of the teachers using Keller Plan meant that to publish a comparison significantly favouring lecture over Keller Plan would be publishing a failure, admitting a waste of time and effort, and challenging a vocal group of enthusiastic teachers. In this situation, selective publication is inevitable, and Kulik et al and Traveggia's conclusions are open to question.

Secondly, I would argue that the potential influence of "Hawthorne effects" in terms of only the novelty of the course was too narrow. Above I argued that differential commitment of the teacher qua evaluator influenced published evaluation results. Differential commitment of the teacher qua teacher is also possible, for even in the long established Keller Plan courses described by Kulik et al, the commitment of the teacher towards
the Keller Plan in particular, (and possibly teaching in general), must be very great. This is manifestly not true in the case of the lecture courses which the Keller Plans are compared with, for it is common experience that few long-established lecture courses are taught by this method because of the active commitment of the teacher to lectures and lecturing. Thus the differential effects of the teacher qua teacher will continue to flaw individual study course versus lecture course comparisons until such time as there are teachers running individual study courses with no more commitment to them than their counterparts running lecture courses. And it is debatable whether such a time should ever be reached. By restricting teacher differences to the effects of novelty (under the umbrella of the Hawthorne effect), Kulik et al ignore the confounding effect of continuing teacher commitment.

I suggest that the above flaws may explain the difference between the findings of Kulik et al and Traveggia on the one hand, and Stice on the other, on the subject of student learning. Stice's project reported all of the individual study courses it initiated (not just the successes), employed an independent evaluation team (who had little to gain from results favouring Keller Plan), and, due to the institutional enthusiasm which it generated for the method, may well have started to involve teachers without the commitment which characterises those innovating "alone" in an institution. Therefore it is concluded that Stice's picture of variation (admittedly favouring individual study quite strongly) is more accurate than the uniform superiority presented by Traveggia or Kulik et al.

Two further conclusions can also be made. Firstly, the difficulty in constructing a valid comparative evaluation in this area is considerable; large research teams with substantial budgets have been unable to achieve this, and overcoming the confounding effects of differential commitment may actually be undesirable. This both provides an indication as to where the research to be reported should, and should not concentrate, and is also a symptom of a broader change in the philosophy of educational evaluation discussed in part II of this chapter. A second, and more tentative conclusion is that rather different
studies of student learning on individual study courses might be productive; some examples would be learning effects on weak students (Elton et al (1973)), or students taking Keller Plan courses to reinforce a lecture series (Allen (1977)).

2.5 Other Topics of Evaluation

Although student reactions to individual study courses, and their learning on them have been the two most researched areas, the evaluation reports in question share other issues. Longer-term retention has been mentioned briefly above. Briefly, Kulik et al (1976) state:

"We located 9 studies investigating retention over intervals ranging from 3 weeks to 15 months. In each of the studies, the PSI students performed better on a follow-up examination than students from lecture courses, and in each study the difference between groups reached statistical significance. If anything final examination comparisons underestimated the magnitude of PSI's effect".

The independent evaluation project reported by Stice followed 7 experimental/control group pairs into later dependent courses to assess retention (although this might more correctly be termed transfer). On none of the later examinations did the students who had completed individual study courses score significantly higher nor lower. Although in this case Stice's project is somewhat flawed methodologically, it does suggest that uniformly positive results may again have arisen from selective publication.

One hypothesis for the various improvements in learning is that individual study courses encourage or demand extra time expenditure on the part of students. Boud, Bridge and Willoughby state:

"The evidence on the amount of time taken by self-paced courses compared with traditional ones is somewhat contradictory. Many of the students in the self-paced courses evaluated by the authors felt that the courses took up more time than if they had been given by lectures. But these feelings were not substantiated by the actual number of hours per week that the students reported spending on the self-paced courses compared with their others".
This early report, and the first two by Kulik et al were unable to go further than this, being based solely upon subjective student recall. Later, Kulik et al were able to add rather less subjective data, including: a study which required students on an experimental/control group pair of courses to work in a resource centre where their time expenditure could be monitored directly; and also an analysis of students weekly recollections of time expenditure. Stice reported a number of investigations taking the latter format, and Kulik et al (1976) summarised the outcome as follows:

"When students report study times frequently during the course rather than once at the end of a semester, their reports agree ..... that students spend about the same amount of time on lecture and PSI courses".

The resource centre experiment confirmed this view, and the contradiction remains. Boud, Bridge and Willoughby conclude:

"The added concentration and effort that a student is likely to spend in mastering a PSI unit, compared with taking lecture notes, say, perhaps causes students to feel that they spend more time, whereas it is rather that, in any given period of time, they work more concentratedly".

Again the background literature has limitations which compound the issue. On this occasion it provides no indication as to the exact workload and workpattern of students doing individual study courses, except in the artificial situation where they are required to study exclusively in a resource centre.

To complete this section there are two areas, the costs of running individual study courses, and the identification of their critical features, which are significant. Of the evaluations under study, only Stice and Boud, Bridge and Willoughby discuss costs, and it has received scant coverage generally, (a notable exception is Black and Boud (1977)). Most authors have isolated two elements for costing, the first is preparation and production. The U.K. authors state "This
involves more staff time, secretarial assistance and duplicating costs than traditional courses". All the courses in Stice's evaluation project substantiated this, particularly those which included non-print media as an alternative to the study guide or textbook.

The second element is the cost of teaching the course once prepared. This depends "... on the size of the class, and the status of the tutors. For example, if the class is so small (1-12 students) that the teacher alone can act as tutor, or if undergraduates can act as (unpaid) supplementary tutors, then the (running) costs may be the same as a traditional course". (Boud, Bridge and Willoughby (1975)). Otherwise, extra tutorial help will be required, and will have to be paid for. Undergraduate tutors are rare in the U.K. (as the survey in chapter 3 will show), and are never paid; they are more common in the U.S.A. and are often paid. Stice found in each case that individual study courses cost more to set up and run than the conventional courses they had replaced. Where proctors were unpaid and the course had run for some time, the running costs were comparable but preparation costs were certainly not recovered.

As well as being rare, the analysis of costing on individual study courses has also been rather unsophisticated. One shortcoming has been the nature of the comparisons, for it is unrealistic to assume that the traditional course is a uniform concept requiring similar resources irrespective of context. Also protagonists of the Keller Plan have argued that the method will always be at a disadvantage where institutions are geared to the lecture system, and point to potential savings in, for example, classroom space. (In practice Stice found the opposite "... the needs for extra space or specialized facilities have been continuing since the project began").

The last area which has received significant evaluation were variants on the Keller Plan, in part attempting to determine the "essential features of the method" Kulik et al (1976). The later two papers by Kulik et al, and that by Traveggia analyse the effectiveness of different features of individual study on the basis of published experimental research on student learning.
Their conclusions are almost totally opposite. Kulik et al (1976) state that the Keller Plan....

"..... seems to work well because it involves (1) small units of work; (2) immediate and specific feedback at every step; and (3) a requirement of mastery at every step. Other features seem to be less crucial: interactions with proctors, self-pacing, and absence of regular lectures". 

Traveggia states:

"The existing research literature suggests that three of the five features probably account for the superiority ....: the go-at-your-own-pace feature; the unit perfection requirement; and the use-of-proctors feature".

Thus although both authors conclude that progressive mastery is essential, the two other features selected by Traveggia are regarded by Kulik et al to be marginal at best; they state:

"..... there is no good evidence that the proctor makes a distinct contribution to PSI's educational effectiveness".

(and cite two studies in which self-marking proved as successful as undergraduate proctor marking). Further they described four studies which show that "there is no difference on final examination performance or student evaluations for teacher-paced and self-paced groups".

Judging between the two is difficult, for they are based on different things. The explanation by Kulik et al is based upon approximately 15 published evaluations involving "addition" or "subtraction" of components of the instructional system. Conversely, and somewhat uncharacteristically, Traveggia cites no studies providing evidence for his analysis, but relies upon elaborating this theoretical model:

"..... the go-at-your-own-pace feature ..... is somewhat misleading. A more appropriate designation would be monitored pacing or forced pacing ..... This eliminates the cramming option and reduces student discretion in scheduling their own studies. Students are 'forced' not to procrastinate ..... What this means for the average student is that the courses tend to be considerably more difficult than conventional courses ..... students not
only spend more time studying than students in conventional courses, but they also evaluate them as being 'a great deal more work'. Thus, at least part of the explanation for the superiority of PSI courses lies in the simple fact that students in these courses are forced to work harder than students in conventionally taught courses. The use-of-proctors feature makes it possible to monitor closely the progress of each student through PSI courses, with the result that lower ability students and/or students who 'fall behind' can be identified readily and given extra help. Finally, a third and equally important part of the explanation for the superiority (of these courses) may be that they have higher drop rates, particularly among lower ability students.

Traveggia summarised the argument in the following diagram:

```
"Unit-Perfection" Feature          "Use of Proctors" Feature
       ↓                              ↓
"Forced-Pacing" Feature  ← Monitored Progression
                  ↓
Students work harder on PSI courses ← Higher drop rates in PSI courses

Superiority of PSI over conventional courses
```

Diagram 21/1

The judgment between Kulik et al and Traveggia hangs upon this argument. Traveggia's viewpoint on pacing is new to the literature, the claim that students at least feel they work harder is also clearly valid, and the monitorial aspect of the tutoring (or "proctoring") system has not always been appreciated. However, the weight of actual evidence questions both assertions in the third row of the model. Kulik et al, Stice, and Boud, Bridge and Willoughby find that the best of the range of
(admittedly inadequate) workload studies do not confirm that
students work harder on Keller Plan courses, and Kulik et al
carefully omitted from their reviews experimental/control group
pairs in which differential "drop rates" occurred, yet they
still found the method superior. All in all Traveggia's model
can be found lacking compared with Kulik et al. The conclusions
to this part of the literature survey will examine why it is
important to determine key features of, for example, Keller's
initial plans.

2.6 Conclusions to Part I

Part I of the literature survey on individual study courses
is necessarily rather long, for the range of background publica-
tions is large, and yet few attempts have been made to gather them
together. In some respects the previous section on significant
evaluations of individual study courses serves as a conclusion
and it reviews data on the "state of the art" in this area.

The literature indicates areas in which research is needed
in two major categories. The first and most straightforward are
a number of detailed questions quite clearly shown up. The
unresolved question of student workload is one example, and the
effectiveness of various elements of Keller's plan is another.
The research reported in chapter 3 to 6 tackles some of these
questions. The second research area indicated is more funda-
mental. It has to do with the need for an examination of the
kinds of procedures appropriate for the evaluation, both by
teachers and by outside researchers, of individual study courses
(and indeed possibly other innovations in higher education).
Developments in teacher and researcher evaluations are indicated
separately. The former springs from the keenness amongst
teachers to evaluate their teaching, often for the first time,
when they take to teaching by individual study. The novelty of
the task, and the inadequacy of some of the results point to
the need to develop a means of guidance for teachers evaluating
their own courses. But a further reason is more critical;
individual study courses are neither trouble-free nor familiar for teachers nor students, and there have been some unfortunate failures. In this situation formative evaluations which formalise and systematise feedback from student to teacher represent a potentially valuable means of "trouble shooting".

Rather different developments in researcher-evaluations are indicated, in part by the inability of the work so far reviewed to fully cater for the needs and the values of the teachers or "clients" they are intended to serve. Most of this research is from the U.S.A., and much of this was conducted in the framework of psychometric evaluation, (discussed in part II of this chapter). But the comments of several of the main "customers" of individual study course research such as Keller, Sherman, and Stice illustrate the need for evaluation research:

"..... I have often felt disturbed when teachers or researchers have suggested that the adoption of this system should depend upon a statistically significant difference in the performance of 'equated' groups of students on some test of subject-matter mastery. First, because of the difficulty in making such comparisons; secondly, because their outcome is commonly ignored unless the differences are so great as to have been apparent to the 'naked eye' well before the calculations; and, finally, because there are other, possibly more important, variables that ought to be considered". (Keller in Stice (1976)).

"..... Even the best of current standard analytic procedures are not sufficient or appropriate to the analysis of the total effect of individualized teaching methods". (Sherman in Stice (1976)).

And Stice's disenchantment is particularly clear:

"I no longer am bothered that our statistical data do not allow me to make an unequivocal statement that PSI is a superior method of instruction". (Stice (1976)).

These three were reactions to the data produced by Stice's project (described in 2.6 above) indicating that the Keller Plan may not have broken the "Teaching-Learning Paradox" (Dubin and Travaggia (1968)).
Elton (1973) expressed similar views, but also pointed towards a solution:

"It seems natural to evaluate an innovation in terms of learning effectiveness in comparison with a control group using traditional methods, but the literature is full of attempts at such evaluations which lead to no or minimal differences. And yet we often feel it in our bones that there are beneficial changes of a less tangible kind, if we could only pin them down. This feeling is far too general to be ascribed solely to the euphoria of the innovator, and it is therefore fortunate that evaluation methods more suitable for detecting such changes than the orthodox 'agriculture-botany' ones are being developed".

Thus as well as researching an educational innovation, the problems of evaluation by teachers, and answering certain questions specific to individual study courses, the literature also indicates a need for research into the process of curriculum evaluation.

The debate described in 2.3 on the essential elements of individual study courses, and the debate on the aims such courses can best achieve (Black (1976)) point to a further question, on whether there is developing a theory of individualized learning which may take over from a predominance of research into teaching method. Section 2.5 left the question of the essential elements of Keller Plan open; the findings of Kulik et al (1976) and Traveggia (1976) being almost directly opposite. A useful postscript to this is the position taken by Keller (1974). Typically, he tackles the question by reference to the reinforcers built into his plan, and these he lists as: getting ahead unit by unit, progressive mastery, small work units, self-pacing, immediate grading, and the tutorial contact with undergraduate proctors. In other words he sits on the fence by stating that each of the features of his plan are essential with the exception of the idea of "motivational lectures" (and chapter 3 will show that this latter concept has been widely dropped in the U.K.). Part I of this chapter has shown that users of Keller's plan have not shared his concern for
preserving each of its features intact.

However, discussion, particularly in the U.S. literature on variants on original schemes by Keller, Postlethwait or others miss an important point. This is the fact that different variants may achieve different aims, rather than the same aims to a better or worse degree. Thus: teacher-pacing may achieve interdependency amongst students where student-pacing aims at independence (Elton 1977); discussion or "non-mastery" tests may provide overview and integration whereas mastery testing may result in more mechanistic learning (Davis 1976); and using class periods for group work rather than one-to-one tutoring may reduce tutor-student contact, but save the teacher's time, and achieve an important aim by providing larger groups of students with individual study courses without extra cost.

It would be unjust to conclude this review of the literature on individual study courses on a note of criticism. Much has been achieved. If there appears to be a leveling off of papers about the Keller Plan itself, this is more than balanced by the diversification and modification it has prompted. The original schemes and their proponents have shown a growing number of teachers in higher education that there is an alternative to the lecture, and experience has produced guidelines which take much of the risk out of innovating in this way.
2.7 Introduction

Surveying the literature on course evaluation for part II of this chapter presents fewer problems than were encountered in part I, where the difficulty was in bringing together for the first time in one review a very large number of publications on individual study courses. Evaluation in higher education presented a similar problem until quite recently. However, since 1974, a number of authors have drawn together the literature on educational evaluation generally, and even the number relating this to the specific context of higher education in the United Kingdom is growing. By using reviews such as those by Tawney (1976), Hamilton (1976), and Boud (1974) as a springboard, this part of the literature survey can move more quickly into areas crucial to the research to be reported subsequently. Thus part II is less descriptive and rather shorter than part I.

There are three main sections in part II. Section 2.9 is on the development of evaluation as a field of study. Starting from the work of Tyler (1949), who saw evaluation in terms of the achievement of objectives, it goes on to examine comparative evaluation and recent moves away from the dominance of psychometrics in curriculum evaluation. Section 2.10 analyses the current literature by means of "dimensions" for understanding and examining evaluations. Lastly, section 2.11 considers the literature which has specific relevance to formative evaluation in higher education.

2.8 Résumé: the Development of Evaluation as a Field of Study

The literature on course evaluation has a short although prolific history, but a longer past. For although it was implicit in many of the schemes of systematic curriculum development
which formed one basis of educational technology, the recognition that it is potentially a discrete field of study, and one which is essentially different although complementary to student assessment, came more recently. This section looks briefly at the curriculum development models which include course evaluation, and in more depth at specific approaches to evaluation.

Authors have by no means been in agreement as to the proper remit of the evaluator or the definition of curriculum evaluation. The following is a working definition put forward to bound the discussion:

"Curriculum evaluation is the collection and provision of evidence, on the basis of which decisions can be taken about the feasibility, effectiveness, and educational value of curricula". (Tawney (1976)).

As this section proceeds it will be clear that different authors interpret this in different ways, and their own definitions will be shown to reflect this.

2.8 1. Evaluation in Systems of Curriculum Development

Amongst the first educators to consider educational evaluation have been the curriculum developers, in particular Ralph Tyler (1949) who is regarded as the founder of systematic curriculum development. The approach which he developed in the 1940's contained evaluation as an integral part. His successors in curriculum development such as Lehmann (1968), and later Stenhouse (1975) and Rowntree (1974) in this country continued this emphasis. Although a worthwhile starting point, the earlier approaches will be shown to be lacking in the present context, and the later approaches provide a link with curriculum evaluation models.

Tyler's well-known starting point for curriculum development was a series of four questions:

"(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 organized?

(4) How can we determine whether these purposes are being attained?" (Tyler (1949)).

Although Tyler did not use the term (here) the emphasis was upon defining objectives, matching them to available learning experiences, and then testing the extent to which the objectives had been achieved. For him, stage 4 was the definition of evaluation, and a measure of how far the purposes (or objectives) were being achieved provided the criteria both for judging the attainment of the students (i.e. assessment), and for judging the value of the course (i.e. evaluation). Tyler was the first to define evaluation in terms of the congruence between teacher intents and learner outcomes, and this "objectives testing" approach to evaluation held sway for thirty years. (One check on the interrelationship between educational objectives and approaches to evaluation is particularly clear. The period of questioning the dominance of teacher objectives characterised by the work of MacDonald-Ross (1973) and Stenhouse (1975) coincides closely with that of the questioning of "conventional" systems of evaluation characterised by Stake (1967) and Parlett and Hamilton (1972)).

Thus Tyler's concept of evaluation depends upon the possibility of measuring the outcomes of instruction rather than the means by which they are attained. Section 2.9.2 will show that this is neither as feasible nor as desirable as it appears at first sight.

Tyler's work was built upon by others, and found particular application in the design of training courses. Mager (1962), Popham (1970), and Lehmann (1968) are just a small sample of the authors who placed different emphasis on the interpretation. For some time the trend was towards increasingly detailed pre-specification, and the need for defining educational objectives in terms of observable and measurable behaviour. Bloom and his co-workers produced taxonomies of educational objectives which were the second cornerstone of systematic course design and evaluation (Bloom (1958), and Krathwohl et al (1964)).
"Course objectives represent a clear statement of instructional intent, and are written in any form necessary to clarify that intent. In practice, you will need to have at least twice as many statements as you have tasks. Unless we know precisely where we are going, we might wind up someplace else and not even know it".

(Mager and Beach (1967)).

Later authors on curriculum development saw evaluation playing a wider and more diverse role, going further than only judging the achievement of the teachers' intents. For example in 1974, evaluation played the following part in curriculum development:

"Having designed a learning experience and exposed the student to it, we must ask: 'Did it do any good? What effects did it have?' Evaluation is the means by which we systematically collect and analyse information about the results of a student's encounter with a learning experience. We wish to know which objectives have been achieved, and which not; and what unforeseen results (beneficial or disastrous) have also materialized? ....... Hopefully, the insights gained from evaluation will help us improve our teaching (or enhance the learning) not just for the present students but for the future students also". (Rowntree (1974)).

Thus curriculum developers were changing their view of evaluation to cover, in this example, unintended outcomes, and evaluation for improvement within and between courses.

2.8 2. **Beyond Evaluation as Objectives Testing; First, Comparative Evaluation**

Evaluation in the Tylerian sense of objectives testing was not the earliest widespread approach in the United Kingdom, where evaluation, and educational research generally has a long history of comparative experimentation. Hamilton (1976) said:

"Perhaps the most pervasive influence in curriculum evaluation has been exerted by a research tradition based on the theory and methods of mental testing, and 'field experimentation',
of the kind predominantly used by agricultural botanists". 
In this approach:
"Students - rather like plant crops - are given pre-tests 
(the seedlings are weighed or measured) then submitted 
to different experiences (treatment conditions). Sub­ 
sequently, after a period of time, their attainment 
(growth or yield) is measured to indicate the relative 
efficiency of the methods (fertilizers) used".

(Parlett and Hamilton (1972)).

Many comparative evaluations took place in the 1950's and 1960's. In the United States of America these ran parallel 
with the objectives testing evaluations, whereas in the United 
Kingdom they predated and predominated over objectives testing. 
Particularly well known comparative evaluations in United 
Kingdom school teaching included an extensive comparison of 
streaming versus non-streaming (BarkerLunn (1970)) and a com­ 
parison of the effectiveness of a new initial teaching alphabet 
(i.t.a.) with traditional forms (Downing (1967)). Additionally, 
in the United States of America, tertiary level education was 
subjected to comparative evaluation. A review of these studies 
was described in part I of this chapter, and the conclusions 
of Dubin and Traveggia (1968) were that they had failed to show 
significant differences between a range of different teaching 
methods. The outcome of comparative evaluation in the United 
Kingdom has been the same; they have been "unable to furnish 
the kind of categorical answer that is expected of them" 
(Hamilton (1976)).

Although comparative evaluations differ from objectives 
testing evaluations both by not being a part of the curriculum 
development process, and by facing the possibility of alternative 
teaching methods, there are considerable similarities between 
the two. Comparative evaluation is also concerned with testing 
prespecified outcomes (although the evaluator rather than the 
curriculum developer may now do the prespecifying), and this 
takes place to the exclusion of unexpected outcomes. Further, 
comparative evaluations do not study differences in processes 
of education beyond ensuring, (usually in the sampling and
statistical analysis), that such differences are constant and uniform. This control of different procedures, and the need to establish the difference between products, has kept comparative evaluation a field for educational testers and psychometricians. Consequently comparative evaluation has been called "psychometric" evaluation, although this term is also used to cover objectives testing evaluations by some authors (e.g. Howe and Delamont (1974)).

2.8 Changing Strategies of Evaluation

Largely because of their drawbacks, both objectives testing and comparative evaluations have been eclipsed by new approaches in the past decade. The main source of the problem with early formal evaluations was in measuring only the predefined outcomes of instruction, and thus it was the focus of the changes that resulted. Newer approaches emphasised study of the teaching-learning process as well as the expected products, and thus they were able to add explanations of the unexpected outcomes of a teaching procedure (either established or innovatory). Additionally in the case of the evaluation of teaching innovations, this broader remit results in a description of the system in practice. The "gross oversimplification of intended consequences" (Hamilton (1976)) was perhaps the key source of dissent and change, for experience of innovations in the 1960's showed that their "constituent elements are emphasised or de-emphasised expanded or truncated as teachers, administrators, technicians and students interpret and reinterpret the instructional system for the particular setting. In practice, objectives are commonly reordered, redefined, abandoned, or forgotten". (Parlett and Hamilton (1972)).

Other details of the new evaluations, and the procedures which spring from them are direct consequences of the emphasis on processes as well as products. Hamilton summarised the characteristics as follows:

"Compared with the classical models, they are more extensive (not merely restricted to test data), naturalistic (based on programme activity rather than programme intent), and
adaptable (not constrained by preordinate designs).
Typically, they aim to provide information rather than
judgment by (1) featuring field studies which portray
the innovation in the context of recognisable reality;
(2) documenting a broad spectrum of phenomena, judg-
ments, and responses; (3) reporting the study in a form
appropriate to the audiences seeking information".

It should be clear from this summary that the change of
emphasis prompted by the "new wave" evaluators was primarily
theoretical. Perhaps inevitably, this has preceded a detailed
analysis of the practical, day-to-day tasks of the newer
evaluations by some considerable time.

It would be unfortunate if the above overview gave the
impression that changing strategies for evaluation came about
easily or as a result of uniform development. The following
paragraphs examine the work of individuals who pressed for
change, not always in the same direction.

In 1963 Cronbach stated:
"I am becoming convinced that some techniques and habits
of thought of the evaluation specialist are ill suited
to current curriculum studies ..... we must depart from
the familiar doctrines and rituals of the testing game".

In this paper he recognised "three types of decisions for
which evaluation is used: (1) course improvement, (2) decisions
about individuals, and (3) administrative regulation". Category
3 concerned "judging how good" is the school, the innovatory
project, or the individual teacher. Scriven (1967) later termed
category 3 "Summative Evaluation", and 1 "Formative Evaluation"
whilst the second tends to be separated out as student assessment.

In spite of his position as a foremost educational psycho-
metrician in the United States of America, Cronbach's theme was:
"Educational experimentation has been concerned with com-
paring score averages of competing courses. But course
evaluation calls for a description of outcomes. This
description should be made on the broadest possible scale,
even at the sacrifice of superficial fairness and precision".
And therefore he stated the object of evaluation as:
"To uncover durable relationships - those appropriate for guiding future programmes".

His definition of evaluation being:
"The collection and use of information to make decisions about an educational programme".

This first questioning of objectives-based evaluation was built upon by Stake in 1967 and Hastings in 1969. Hastings pointed to the fact that evaluators had attended to their "kith and kin" in disciplines such as psychology and agronomy to the exclusion of economics, sociology or history. In doing so he suggested specialisms within which Cronbach's ideas about more eclectic evaluation strategies might grow.

In 1967 Robert Stake made perhaps the major contribution of the 1960's when he described an analytical framework which showed what the new evaluators might actually concentrate upon and do. Again he diagnosed the problem in his own way:
"Today's educator may rely little upon formal evaluation because its answers have seldom been to questions he is asking". (Stake (1967)).

Stake points out that two roles of evaluation, description and judgment, must both be present to overcome this. As a result, he put forward a model of evaluation based on a description and a judgment matrix.

These matrices were a way of showing how evaluations could go beyond the preordinate "objectives testing" approach. Each matrix contains three stages in the process of course development: antecedents prior to the course, transactions during the course, and outcomes of the course. The description matrix takes as its other dimension intents and observations; terminology chosen to emphasise that as well as the teachers' objectives, hopes, and fears, those of students or other interested parties should also be attended to in planning the evaluation and describing the outcomes of a course. The judgment matrix takes as its second dimension: educational standards required by different reference groups, and judgments made about the achievement of these standards. Finally Stake added the rationale of the
course or programme as a separate element in the model intended to provide one basis for evaluating the range of intents. He represented the model diagrammatically:

<table>
<thead>
<tr>
<th>INTENTS OBSERVATIONS</th>
<th>STANDARDS JUDGMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTECEDENTS</td>
<td></td>
</tr>
<tr>
<td>TRANSACTIONS</td>
<td></td>
</tr>
<tr>
<td>OUTCOMES</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2II/1 Stake's Model of Evaluation**

Analysis of this model provides Stake with approaches to evaluation which go beyond objectives testing. Looking across the diagram they include seeking the answers to questions such as: Do the "course objectives" category of intended transactions and outcomes naturally spring from the rationale of the programme? How far is there congruence between the intended antecedents, transactions, and outcomes, and those actually observed? Is there logical contingency between the intended antecedents, transactions, and outcomes, and how far is this contingency observable in practice? How does the description matrix compare with externally set standards of excellence, or with the description matrix of some alternative course? And finally, what judgements can be made on the basis of these comparisons about the course or programme in question?

Stake's model has rarely been explicitly tested, a characteristic shared with other evaluation models. It is important because it highlighted the complex range of questions which evaluations might tackle, thus sparking off the reappraisal of the role of evaluation.
which can be viewed from this framework.

Perhaps the most significant follow-up to Stake's analysis was the work by Scriven in the United States of America (Scriven (1972) and (1973)) and Parlett and Hamilton in the United Kingdom (Parlett and Hamilton (1972)). Both proposed non-prescriptive evaluation models, and were able to use a growing weight of evidence to criticise preordinate evaluations.

Scriven's treatment of intents for course development was particularly extreme, suggesting "that consideration and evaluation of goals was an unnecessary but also a possible contaminating step ..... and the less the external evaluator hears about the goals of the project, the less 'tunnel-vision' he will develop, the more attention will be paid to looking for actual effects (rather than checking alleged effects)". (Scriven (1973)). Scriven calls his alternative "Goal-Free Evaluation", and requires the evaluator "to refrain completely from any personal discussion of goals with the programme sponsors or staff. The evaluator, perhaps with the help of colleagues and consultants, then is expected to recognise manifest goals and accomplishments of the programme as he works with it in the field ..... what is intended is not important, the programme is a failure if the results are so subtle that they do not penetrate the awareness of an alert evaluator".

However, the notion of evaluators avoiding programme intents has not become widespread. Stake felt that this was because Scriven's goal-free evaluation expects evaluators "..... to be as sensitive, rational, and alert as his designs for evaluation require". (Stake (1974)).

Like Scriven, Parlett and Hamilton put forward a new approach to evaluation which they give a name: Illuminative Evaluation. Again this approach has more to do with overall evaluation of course programmes from the outside than day-to-day evaluation of course details by more involved parties. In the five years since it was first described, Illuminative Evaluation has gained more supporters than perhaps any other non-traditional evaluation procedure, and rather like Keller and Green's founding papers on the Keller plan (Keller (1963), Green (1971)), Parlett and Hamilton's paper has been widely read and reprinted (for example in Tawney (1976)).
A brief description of Illuminative Evaluation will bring this section to a close, and the history of educational evaluation up to date.

Parlett and Hamilton opened by criticising preordinate and comparative evaluations (which they ascribe to the "agricultural botany paradigm") on several counts. Three were particularly relevant:

1. By producing a single "objective truth" equally relevant to all parties interested in the evaluation results, the approach fails to address the varied range of needs of teachers, researchers, and students, for example.

2. Only data which can be quantified tend to be gathered, to the exclusion of subjective, anecdotal, or impressionistic data.

3. Attempts to produce statistical generalization fail to study unusual effects, or local perturbations which have great potential significance.

The alternative presented by Parlett and Hamilton, Illuminative Evaluation, is said to owe more to the social-anthropology paradigm. It "... takes account of the wider contexts in which educational innovations function. Its primary concern is with description and interpretation rather than measurement and prediction". Two foci of an Illuminative Evaluation are the instructional system and the "learning milieu". Parlett and Hamilton suggest that the first should be treated as a catalogue description. Unlike Scriven's goal-free approach, the evaluator examines the plans and goals of the originators, but adds to these the goals of teachers, students and outsiders when considering the subjects for study. Descriptions of the learning milieu ("the social-psychological and material environment in which students and teachers work together") and the way this interacts with the instructional system contribute further to the approach, and only then are they followed by "progressive focusing and concentrated attention to the emerging issues".

Parlett and Hamilton suggest new priorities as well as new foci. The most basic springs from their definition of the purpose of evaluation: "to contribute to decision making". However, in practice,
some evaluations may be used to delay troublesome decisions, or to window-dress a policy already formulated ...." This recognition of a variety of different uses for evaluation results in research priorities and reporting procedures being fitted to the range of different audiences.

Illuminative Evaluation was significant because it was comprehensive. In terms of Stake's model, it took the position of accepting the rationale and the intended antecedents, transactions, and outcomes of a course as just one input into evaluation planning, to be added to by observed antecedents and more particularly by the transactions in the "learning milieu". The observed outcomes were to be portrayed descriptively and qualitatively, according to the needs of different audiences. However, in this lies a fundamental disagreement between the United Kingdom and the United States of America authors, for whilst Stake includes a judgement matrix, and suggests means by which evaluators may provide judgments of educational projects, Parlett and Hamilton (1972) see this as problematic:

"A decision (judgment) based on one group's evaluative criteria would, almost certainly, be disputed by other groups with different priorities .... The investigator does not make decisions .... (but) provides a comprehensive understanding of the reality (or realities) surrounding the project".

This disagreement about the evaluator's responsibility for judgment represents just one of the ranges of discussion which are described in the next section. The intention is to suggest reasons why either position might be appropriate in different circumstances, rather than take one or other point of view.

2.9 The Dimensions of Evaluation: Many Ways of Looking at the Task

The preceding sections describe what might be thought of as the changing philosophies of evaluation, and they show how the emphasis has shifted from objectives testing to a more eclectic approach. Although this is vital to a discussion of evaluation, it falls a long way short of describing the task of the evaluator.
However, since evaluation emerged from Tyler's curriculum development systems, much has been written about how, when, by whom, and for whom evaluation should take place.

This section suggests a new way of looking at such literature. It is presented in terms of a number of dimensions along which the task of evaluation can be examined and decided upon. The analysis is somewhat tentative at this stage, and there is scope for future development. It has two advantages: it portrays evaluation as a continuum of different theoretical and practical emphases, rather than a series of separate and opposing camps; and it does encompass the significant areas of debate about evaluation today.

Six dimensions have been identified as relevant to the research; three of these appear fundamental and are dealt with first:

- The GOAL POSITION of evaluation, lying on a range from goal judgment, through goal acceptance and goal consciousness, to goal-free evaluation.

- The REMIT of evaluation ranging from descriptive to prescriptive (or judgmental).

- The TIMING of evaluation, lying on a range from evaluation during course development aiming at course improvement, to final evaluation for course appraisal.

The place of prespecified GOALS in evaluation is the first dimension, and approaches have been suggested which place a variety of emphasis upon this. The range might be represented as follows:

<table>
<thead>
<tr>
<th>STAGE 1</th>
<th>STAGE 2</th>
<th>STAGE 3</th>
<th>STAGE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation as appraisal of the value or appropriateness of the prespecified goals</td>
<td>Evaluation solely concerned with measuring the achievement of prespecified goals</td>
<td>Evaluations concerned with both prespecified and other more diverse, or unexpected outcomes</td>
<td>Evaluations conducted in ignorance of prespecified goals</td>
</tr>
</tbody>
</table>

Figure 2II/2
Stage 1 is commonly known as validation of the aims of the course (Tyler (1949)), and as such is an important component of Tylerian curriculum development models of evaluation, since it addresses the question "what educational purposes should the school seek to attain?" As a possible focus for course evaluation, validation has been widely ignored, but recently Hamilton (1977) has redressed the balance, and suggested that it is a vital feature of "strong" evaluation:

"Weak Evaluation is the scrutiny of a given educational practice which takes account of the values of those who share an interest in its activities and outcomes. Strong Evaluation is the scrutiny of a given curriculum practice and the scrutiny of the values of those who share an interest in its activities and outcomes. Thus, strong evaluation relates not only to the study of educational practices, but also to the examination of educational ideas and criteria".

By extending preordinate evaluation (stage 2) back into validation, (stage 1) Hamilton thus provides the logical corollary to his development, with Parlett of Illuminative Evaluation towards stage 3. Stages 2 and 3 represent the "intents" and "observations" stages of Stake's description matrix, stage 2 covering both evaluation via objectives testing, and via comparison. Stage 3 is the position taken by many of the "new wave" evaluators. And stage 4 represents Scriven's Goal-Free evaluation.

The REMIT dimension of evaluation is both fundamental and controversial. Again Stake's model shows the range of possibilities, with both a description and a judgment matrix. The judgment matrix emphasises the possibility of either absolute or relative judgments. In practice, evaluators have construed this dimension as a question of how far they should interpret the descriptive phase of their work:

"There were questions, too, of how evaluation studies should be reported. Should they constitute a 'display', in 'raw' form, of the range of different opinions, results from questionnaires, and so on? or, alternatively, should the
report distil, summarise, organise, and interpret the data for its different audiences?" (MacDonnald and Parlett (1972)).

The arguments advocating the descriptive approach suggest that it is the client of the evaluation's job to evaluate in the literal sense of the word, and the evaluator's to provide the reader with information which he may or may not take account of in making a decision. However, Stake (1974) (and also Scriven (1967)), would reply that "straight narrative reporting" is a misnomer anyway; that interpretation is inevitable and desirable; and that the evaluator is best suited to doing it. Stake, quoting Scriven stated:

"Description is one thing, judgment is another. Most evaluation specialists have chosen not to judge .... Scriven's position is that there is no evaluation until judgment is passed, and by his reckoning the evaluator is best qualified to judge".

Like the other dimensions there is no "correct" position. The non-judgmentalists do make clear one important means of deciding upon a position; the needs of the clients of evaluation. Taylor and Maguire (1966) pointed out five groups of possible clients: "spokesmen for society at large, subject-matter experts, teachers, parents and students themselves". The difficulty of providing judgments acceptable to most or all of these groups is considerable. Conversely Kemmis (1976) outlined two occasions where a descriptive "portrayal" is desirable:

"The first is when an interested audience cannot directly experience the programme, and cannot accurately judge it without some 'surrogate experience' of it. The second is when participants in the programme are too closely involved in its functioning to be able to step back from it and see it as a whole ...."

Thirdly, there is the dimension of evaluation TIMING. The dimension here goes from "evaluation for course improvement" (Cronbach (1963)) to evaluation for "final appraisal". In 1967 Scriven entitled these formative and summative evaluation respectively. Formative evaluation generally takes place where there is some commitment to development and change on the part of teachers
and evaluators, often during the formation of an innovative teaching procedure. It is generally concerned with gathering interim data about the progress of the course. This is then treated as feedback for the teacher who may make changes on the basis of it. Summative evaluation takes place on a finished or nearly finished product, without necessarily requiring the involvement of the teacher, his sanction, or commitment to change. Its purpose is to assist decisions about the overall worth of the procedure as it stands. (Whether it provides absolute or relative judgments as well as a description, and whether it is couched in terms of prespecified goals, depends upon the position taken in the two previous dimensions.)

However, in practice the categorization of evaluation as either formative or summative is not always helpful. For example, summative evaluations which locate major problems may lead to course improvement, and conversely, the progressive feedback produced by formative evaluation may mount up to produce an overall assessment of the procedure developed. In surveying the field of evaluation for its contribution to developing computer aided instruction, Howe and Delamont (1974) describe another shortcoming of this categorisation:

"In effect, formative evaluation is defined as providing 'insights' for the developers, while summative treats the innovation as a completed 'product' and studies how it works when implemented. However...... acceptance of 'the innovation' as a reified entity is not a good evaluative strategy in many cases, and this distinction is not, therefore, very useful...... Just to talk of 'summative' implies the pristine finished product being launched upon an unsuspecting world - a notion which prevents any real understanding of the events during implementation".

There is a slight flaw in this critique, for it is only preordinate summative evaluations which run the risk of treating their subject as a "pristine finished product"; goal conscious or goal-free evaluations have readily avoided this. See, for example, the summative evaluation of the less-than pristine development of Keller plan courses at MIT (Hirshi et al (1975)).
As stated previously, Cronbach's work predated Scriven's formative-summative categorization, and he referred to the two concepts as "evaluation for course improvement", and for "administrative regulation". In both cases he asserted that evaluation procedures need to be more thorough and extensive where the subject is an individual teacher or student under evaluation than when it is a course or a programme.

Thus the goal position, the remit and the timing of evaluation represent three important dimensions for deciding the total picture of evaluation. However, there are many other dimensions. The "by whom?" and "for whom?" questions provide two.

A useful way of looking at who does the evaluation is in terms of their degree of involvement with the subject. The range might be seen as extending from the kind of informal day-to-day evaluation by teachers of their own courses, through teachers evaluating one another's courses, and through outsiders participating in the course and evaluating it, to non-participant evaluation by outsiders. It has been shown that early concepts of evaluation initiated by curriculum developers like Tyler (1949) made evaluation an integral part of the developer or teacher's job. The later emphasis upon psychometric testing put evaluators at the other extreme; outside experts in educational testing. And the new approaches to evaluation have not changed the focus significantly. They are primarily concerned with the evaluator who is a non-participating outsider rather than the teacher himself.

Participant evaluation by outsiders has been somewhat rare despite its popularity amongst sociologists, and anthropologists (Young (1972)). One possible explanation is that summative evaluation may be more credible when generated by non-participants, and formative evaluation more potent when performed by the teacher himself. An important exception is the work on participant evaluation by Howard Becker, initially in examining student culture in medical schools in the late 1950's and early 1960's (Becker et al (1968)). The result was an "analytic field work characteristic of participant observation ..... (which brings
out) the fact that the technique consists of something more than merely immersing oneself in data and 'having insights'". The "new wave" evaluators are increasingly looking to research of this kind, and the result may be greater emphasis upon participant evaluation by outsiders.

All teachers evaluate their courses, albeit in an informal and often piecemeal fashion. So far professional evaluators have made little contribution to this process, nor made use of its results. Chapter 5 of the thesis describes a programme which I initiated to help teachers evaluate their own individual study courses on the basis of little background literature.

Following on from their objection (cited earlier) to the concepts of "formative" and "summative" evaluation for CAI projects, Howe and Delamont (1974) define two different versions of evaluation, one of which has potential relevance to teachers. As with the research to be reported later, they called this version "monitoring" which is:

"Designed to iron out 'bugs' in the innovation ..... Monitoring demands someone close to the project team, ..... its purpose is a servicing device, and its audience is the project-team".

However, in practice CAI projects have tended to include a "monitoring" evaluator on the course team rather than combining the roles of teacher, developer, and evaluator as is necessary and realistic in many situations.

At school level the Ford Teaching Project (Elliott and Adelman (1975), which developed procedures of inquiry/discovery teaching, promoted and formalized teacher evaluations in the formative stages of the project. Teachers recorded classroom events, undertook interview and questionnaire programmes, and developed other means of self-evaluation. However this project is still writing up an appraisal of the teacher evaluations, although numerous interim papers chart their progress (see Elliot and Adelman (1973)).

The evaluation dimension to do with the "user", the "audience", or the "client" of the evaluation also ranges from clients solely within the course or project to wider audiences in the outside world.
The dimension is important because "the evaluation which changes government policy will not be identical with the report that the project director wants to see privately, that the headmasters want to wave at parents, or which the learned journals want to publish". (Howe and Delamont (1974)).

Different authors have placed different emphasis on the various parts of the range. Recently it has been accepted that evaluations must cater for "multiple audiences" (Stenhouse (1975)), and also a growing number of evaluators have proposed versions of evaluation which concentrate on the neglected extreme of teacher-centred evaluations (these include Stake's "Responsive Evaluation" (1972) and Boud's "Supportive Evaluation" (1974)).

Finally, there are dimensions relevant to the subject and scope of the evaluation. The scale of the course, innovation, curriculum or project is one such dimension. Predictably, most professional evaluators have been concerned with larger scale subjects for they tend to be more noticeable or more costly, more controversial, or of more general interest than small individual initiatives by teachers. Thus, for example, Parlett cites the following examples of Illuminative Evaluation:

"A study of Wellesley College in Massachusetts; university examinations at Edinburgh University; science teaching in remote rural schools in the Scottish Highlands and Islands; experimental undergraduate teaching programmes at MIT; and patterns of educational provision for the visually handicapped in England and Wales". (Parlett (1976)).

This is reasonably representative of the subjects taken by evaluators. And again there is an indication that the pendulum of evaluation may have swung too far, in this case, from Tyler's emphasis upon evaluation being part of the curriculum development function (no matter how local), which is undertaken by every teacher.

A further variable in this context is the scope of the evaluation once the subject has been chosen. For example, evaluation of an innovatory undergraduate laboratory course might look no further than the events and accomplishments during the laboratory classes, or it might cast the net wider and assess the
effects on students' study patterns, their attitude to the whole undergraduate course, or their eventual career decisions. Because they focussed upon measuring teacher-objectives, traditional evaluations tended to take a restricted scope, but the results were often confounded or influenced by effects in the wider sphere, (e.g. BarkerLunn (1970)). Conversely, recent approaches have emphasised the "learning milieu" or the "total context", although this may be a consequence of the larger scale of their subject matter.

Finally, Scriven (1967) has made a distinction between "intrinsic" and "payoff" evaluation, and this has been developed by Eraut (1972) who uses the term "performance" instead of payoff. The subject of intrinsic evaluation is the teaching material itself, the curriculum proposals, objectives, methods, and tests. It is an abstract activity involving analysis of these intrinsic features of the course in order to discern its likely planned and unplanned effects. Payoff evaluation examines whether these effects appear in practice, and whether there have been any further unplanned effects.

The purpose of presenting the above dimensions of evaluation was to help clarify the options open to an evaluator, and the decisions that need to be taken when planning an evaluation. Their description has brought out many of the developments in evaluation theory, and shown how the task of evaluators had changed. The dimensions are summarised on table 2.1 It is important to realise however, that many of the dimensions and their corresponding ranges overlap, nor are they necessarily exclusive or opposite.
Table 2.1

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>The GOAL POSITION</td>
<td>Validation — Goal Based Evaluation — Goal Conscious Evaluation — Goal-Free Evaluation</td>
</tr>
<tr>
<td>The REMIT of Evaluation</td>
<td>Descriptive — Judgmental</td>
</tr>
<tr>
<td>The TIMING of Evaluation</td>
<td>Formative — Summative</td>
</tr>
<tr>
<td>The POSITION OF THE EVALUATOR</td>
<td>Teaching the subject of evaluation — Outside the subject, and not participating in it</td>
</tr>
<tr>
<td>The AUDIENCE OF EVALUATION</td>
<td>The teacher himself — the academic community — the outside world</td>
</tr>
<tr>
<td>The SUBJECT OF EVALUATION</td>
<td>Large Scale Subject — Small Scale Subject</td>
</tr>
<tr>
<td></td>
<td>Broad scope of investigation — Narrow scope of investigation</td>
</tr>
<tr>
<td></td>
<td>&quot;Intrinsic&quot; evaluation — &quot;Payoff&quot; evaluation</td>
</tr>
</tbody>
</table>

2.10 Formative Evaluation of Undergraduate Teaching in the United Kingdom

Up to now in this literature review on course evaluation the discussion has been quite broad; being equally relevant to education or training at almost any level. This section describes previous work on evaluation more directly relevant to this thesis. In practice, the amount of relevant work is small, for the majority of evaluation studies have taken place in the United States of America, and at primary and secondary school level.

In 1974 when Boud reviewed the literature in this area he found only two concrete examples of formative evaluation of
undergraduate teaching in the United Kingdom: the work of Eraut (1972) and Dowdswell (1972) in evaluating the Inter-Universities Biology Teaching Project; and the Open University's experiments with pretesting courses prior to general release. The latter procedure has recently been reported to produce worthwhile course improvements (THES (1977)), and this may herald a growth of interest in formative evaluation in an institution which clearly stands to gain a great deal from it.

Since 1974 there has been some growth elsewhere in the United Kingdom. Perhaps the strongest influence has been the American practice of students evaluating lecture courses. This was summarized in 1975 by Colin Flood Page in "Student Evaluation of Teaching: the American Experience", and outcomes include the establishment of a Student Feedback Unit at North East London Polytechnic (Ramsden (1976)). Developments of this kind have tended to concentrate upon developing a feedback questionnaire for use on a wide range of traditional lecture courses. The teacher agrees to the evaluation, and may assist in the data gathering. The evaluator may be both technician (processing the returned questionnaires to produce a summary or report), or a counsellor (helping the teacher to develop his teaching on the basis of the evaluation results).

A second stimulus for evaluation has been the growth of teaching innovation in higher education; the work of curriculum projects, and particularly, the development of computer assisted learning. The following paragraphs examine literature on both influences, but concentrate on the second which is of greater relevance.

Boud was one of the earliest evaluators to develop formative evaluation of undergraduate science teaching, both for doctoral research (1974) and in the Course Evaluation Research Project (1976). At first "the approach was founded on the idea that if a course evaluation instrument was found to be useful in a particular context, then it was likely that with suitable modification it would also prove to be useful in other similar situations". This conception resulted in a rather technological approach to evaluation which included little personal contact or support. As this became
apparent, the goals of the research were modified.

One notable course evaluation instrument was developed in the earlier stages; the Laboratory Aims Questionnaire (described by Boud (1973) and modified for use in curriculum development by O'Connel, Penton and Boud (1977)). The method involved listing many possible aims for the laboratory course on a questionnaire, and asking "staff and students to rate the importance of these aims in the cases (i) of each group's conception of what an ideal course in the subject should be, and (ii) each group's rating of the ..... (existing) course". The extent to which staff and students agree on the aims of an ideal course, and those which an existing course achieves, can be represented graphically. Areas of disagreement between staff and students, or between "ideal" and "existing" courses would then be a focus of attention.†

The aims questionnaire developed by Boud, and semantic differential techniques were used in a variety of traditionally taught courses, in pursuing the original research aims. Often the teachers had "no tangible commitment to change" and these procedures, "which did not take up much time and were not disruptive to the course" were very appropriate ..... "However, little or no change was seen in ..... the courses".

Boud (and the project's research student, Lynette Willoughby) then worked on a different evaluative approach with a different group of teachers; innovators who were just starting to move away from lecture courses or long-established laboratory courses, towards the kinds of individual study courses to be described in this thesis. Boud named the new evaluative approach "Supportive Evaluation". It involved close contact with the teacher and the courses which were monitored and observed in situ. The teachers were involved in defining evaluation problems and procedures, and as a result they were more likely to modify their courses accordingly. The evaluator did not

† It may be noted in passing that this procedure takes a "goal conscious" standpoint on the dimension of evaluation referred to earlier as "Goal Position".
impose his evaluation procedures or priorities, rather:

"Emphasis is placed upon being supportive and developing a concern for the person and understanding the views that are presented".

The evaluation involved "conjoint monitoring", and one aim was that the evaluator, rather than be the essential ingredient, (as in his original notion of evaluation), should withdraw leaving the teacher able to "rationally evaluate his own teaching activity". Boud reported the success of supportive evaluation in helping decisions and change. However, the time-scale of the research meant that the later stages of progressive withdrawal, and evaluation by teachers were not explored. Nonetheless, the background was set for research into evaluation by teachers reported in chapter 5.

Far more recently, in 1977, one other United Kingdom author has contributed to the field of formative evaluation of undergraduate teaching by individual study courses. Watson (1977) reports the evaluation of 1 unit of a BEd course taught by Keller plan, and this might be looked at as a case study in evaluation by teachers. Chapter 5 contains two case studies of evaluations by teachers who used the resource package I developed for this purpose (entitled Monitorkit (Bridge (1975)), a supporting document to the thesis). Some of the problems reported by Watson, such as conducting and analysing teacher-student interviews, and developing a course questionnaire, are tackled in the package.

Another kind of individual study, which is excluded from most of the thesis as it rarely comprises a total course, is computer assisted learning. Its substantial growth and development underwent planned evaluation from the early stages (Hooper (1975)). The UNCAL project was established to perform overall evaluations of the National Development Programme for CAL, and at local level projects have employed staff in formative evaluation. The latter resulted in many individual contributions on the technology of evaluation (see for example McMahon, Anderson and Barton (1977)) and some on more general methodologies such as Howe and Delamont's work, referred to earlier, on "monitoring"
by project staff, and "non-reactive evaluation" by central
programme staff.

The Nuffield Group for Research and Innovation in Higher
Education undertook a three year study evaluating a very broad
range of developments in university and polytechnic teaching
(Hewton et al (1975)). And there have been a number of
illuminative evaluations of university teaching (cited pre­
viously, for example Miller and Parlett (1973)). However,
neither were formative in purpose, being mainly descriptive
research into the state of the art.

Finally, the American practice of students evaluating
lectures has made a start in the United Kingdom. One important
difference is that the United States lecture evaluations often
have little formative purpose; Flood-Page (1975) and Sherman and
Winstead (1975) assert that in the United States of America,
"the type of student evaluations generally used provide almost
no information with enough utility to suggest teacher change".
Instead the uses for student ratings include ..... "providing
course-end feedback for instructors, evaluating teaching com­
petence for promotion purposes, ..... providing the student
body with information for selecting courses, etc.". Some U.S.
universities go beyond this summative purpose, and integrate
evaluation with support and training for teachers (e.g. the
University of Texas at Austin).

In the United Kingdom lecture evaluations are almost
entirely formative. In both countries questionnaires are often
used, and it was quickly seen that formative evaluation required
different, less global questionnaires than are used in the
United States campus "popularity poles", ..... "when the goal
of the instrument is to foster improved instruction, evalua­
tion after the fact in terms of generalities is not helpful,
because instruction consists of specific behaviours". (Sherman
and Winstead (1975)). Thus at HELP Ramsden reported using
questionnaires assessing specific statements such as: "Returns
work promptly", or "Makes a genuine effort to get students in­
volved in discussions". Other attempts to develop the "specific
behaviours" in lecturing have included the work of Casper (1973)
and Carroll (1973), and the application of Kelly's repertory grid (Kelly (1955)). Casper and Carroll replaced the questionnaire with a post-graduate student observer working from a precise checklist of behaviours, and this has been used in the United Kingdom by Hodgson (Elton (1977)). I also used this method in the evaluation of the optional lectures on the particle mechanics course reported in chapter 4.

Finally, in at least two U.K. universities, staff have used a system of "self-help" for lecture evaluation. Black et al at Birmingham University (1976) report co-operation between academic staff in evaluation via observation, questionnaires, interviews, and video-taping. This is proposed by Kennedy (1975) at Edinburgh who sees evaluation of teaching as a "group activity .... (involving) non-destructive testing". Group development and evaluation amongst academics is also underway at Surrey and Sussex Universities. The report by Black et al is particularly relevant to this research since it also describes most of the methods of evaluation by teachers described in Monitorkit, (albeit in a rather different setting). This supports the intrinsic assumption of that part of the research, that university staff can and will evaluate teaching, and they do so from a viewpoint with many advantages.

2.11 Conclusions: The Literature on Course Evaluation, and its Relevance to Individual Study Courses

Part II of this chapter has examined the literature on course evaluation from the time when it first emerged as an integral part of systems for course development pioneered by Tyler. It involved judging the worth of educational objectives, and measuring their achievement. This approach, and evaluation by comparing the achievements of different teaching methods predominated until quite recently. In the 1960's and 1970's the new evaluations appeared, prompted by some noticeable failures of the old. Their proponents included Stake, and Parlett and Hamilton; and they emphasised field research into processes of education as well as
experimental research into products. Each philosophy stops a long way short of describing the kinds of decisions an evaluator must take about his work, so section 2.9 analysed this in terms of "dimensions". Much of the literature quoted in 2.8 and 2.9 came from contexts quite different from the evaluation of an innovation in undergraduate science teaching in the United Kingdom, so finally, 2.10 looked at the small collection of research in this area.

What can be concluded about the relevance of this area of literature to the aims of the research? Firstly, there are reservations about conceiving of evaluation as a requisite and integral step in systematic curriculum development. Part I of the literature survey shows no examples of curriculum development which followed Tyler's precise steps (nor for that matter the more flexible approaches of Rowntree (1974) for example). This view will be further supported in chapter 3 which surveys the development of individual study courses in the United Kingdom, and the picture becomes clear that teachers have (1) not used formalized systems to develop I.S. courses, and (2) not considered course evaluation in the early stages of this development.

Preordinate evaluation models present other problems in this context, to do with objectives. Both part I of this chapter, and chapter 3 and 4 show that teachers' objectives for individual study courses are sometimes ambitious and imprecise, and they often gradually change away from those of the traditional course they replace. The former hinders psychometrics, and the latter tends to invalidate comparisons with lecture courses. The validity of such comparisons is important, and it pervades both parts of this chapter. For if teachers' aims for individual study courses differ as much from those for lecture courses as do those for, say, project work in the laboratory differ from scripted experiments (Ogborn (1977)), then comparisons such as those made by Kulik and his co-workers (1976) or by Traveggia (1976) lose their value. Chapters 3 and 4 investigate these aims in more depth, and discuss how the achievements of individual study courses might be validly assessed.
The ideas embodied in the new evaluations overcome most of these problems, but present some of their own. The most important stems from Stake's comment that goal-free evaluation expects the evaluator "to be as sensitive, rational, and alert as (the) designs for evaluation require". The difficulty is "evaluating" the new evaluations; making the procedures and analyses genuinely open to scrutiny, and ultimately, going beyond a reliance upon the qualities of the evaluator. The chapters which follow all face this problem.

The dimensional analysis of evaluation has yet to stand the test of time and application, but there is one clear generalisation. The TIMING dimension shows that the formative/summative distinction is perhaps arbitrary; evaluations may perform both functions, depending on when (and how) they are initiated. This being so it will be important to examine whether the research to be reported, which was originally conceived of as formative evaluation, benefits from this categorisation, or whether it strays beyond it.

There are, however, a few respects in which the new evaluations have moved so far from Tylerian approaches as to leave gaps, some of which this research tackles. Two aspects of this were mentioned in part II. Firstly, evaluation has been taken out of the hands of teachers, despite the economic reality and potential relevance of teacher evaluations. And secondly, the new evaluations have tended to be large in scale and summative in purpose. Chapter 5 presents two overlapping developments (the resource package "Monitorkit" and a conceptual model of evaluation at the "micro"-scale) which attempt to redress the balance.

Finally, on part II, both Stake's model of evaluation perspectives, and the analysis by dimensions show that any evaluative research must depend on many potentially hidden decisions about what evaluation is, and what an evaluator actually does. The concluding chapter 7 discusses the decisions taken in this research, and their appropriateness to the aims. At this stage it is possible to start the discussion and also perhaps, bring together sections 2.8 and
2.9. Section 2.8 said much about a change in evaluation philosophy, from "preordinate" or "psychometric" approaches to new evaluations described as "responsive" or "illuminative". And 2.9 described a way of looking at the practice of evaluation along dimensions. Putting the two together it is clear that preordinate evaluations take quite different positions on these dimensions as compared to the new evaluations. The former take a "GOAL POSITION" of acceptance and a judgmental "REMIT", the latter emphasises goal freedom, and description. As a result, the two groups have seemed to have little in common, but perhaps as the range of options becomes clearer, the discussion will move on from which is the "best" approach, to what is the most appropriate mixture of dimensions.

Taking chapter 2 as a whole, certain final issues remain. At the end of part I some of the important clients for individual study course evaluation, such as Stice, Keller, Sherman, and Elton diagnosed problems with evaluation. Their prognoses can be examined in the light of part II, bearing in mind that their comments were made when the new evaluations were still little known. Only Sherman saw the future in terms of developing preordinate evaluation via "new techniques of statistical analysis" ..... "even the best current standard analytic techniques are not sufficient ....." (Stice (1976)). Stice, Elton and perhaps surprisingly, Keller disagreed: "because of the difficulty of making ..... comparisons; because their outcome is commonly ignored unless the differences are so great as to have been seen ..... well before calculations .....; and because, there are other, possibly more important variables that ought to be considered". (Keller (1976)). (In part II these "variables" have been called "goals".) And Elton, perhaps nearer to the heart of the changes in evaluation was able to point to new evaluation methods "more suitable (in this instance) than the orthodox 'agriculture-botany' ones". (Elton (1973)). Thus some of the originators of individual study had first hand experience of the evaluation issues discussed in part II, and it will become clear that their majority view has been accepted in this thesis.
A second observation, generalisable between parts I and II is that the literature grew rapidly during the course of the research. Thus some important contributions appeared during or after the research took place. Most notable were the "meta-evaluations" by Kulik and the other researchers reported in 2.5, and the development of a more flexible "educational technology in curriculum development" by, for example, Rowntree and Stenhouse. However, they have not, and probably could not completely resolve the vast diversity of individual study courses, nor the perplexing vagaries of curriculum evaluation. The chapters which follow may help in this task.
CHAPTER 3: THE DEVELOPMENT OF INDIVIDUAL STUDY COURSES IN THE UNITED KINGDOM

3.1 Introduction
3.2 Early Development in the United Kingdom
3.3 The Methodology of the Survey
3.4 Results of the Survey
3.5 Discussion; the Emerging Trends
3.6 Conclusions and a Future Perspective
3.1 Introduction

This chapter describes the way in which a small but growing number of science teachers in higher education have learnt about, adopted, and changed individual study courses. Many of the methods derive from work in the Americas, such as that by Keller, Green, and Postlethwait, described in chapter 2. The basis of the chapter is a survey of individual study courses in the United Kingdom which I carried out in 1975.

There are two other elements to this chapter. Before introducing the survey it is important to set the scene by describing the beginnings of individual study courses in the United Kingdom (section 3.2) which took place between 1970 and 1974. Secondly, there are a great deal more data about many of the courses surveyed than were gathered in the survey programme alone. This is because during the first and second years of the research, I was often directly engaged in their evaluation. These early evaluations are not described fully in this thesis, although their aims, methods, and outcomes are generalised in section 5.2, and each evaluation report is a referenced supporting document. Instead they are drawn upon to illustrate trends and generalisations from the survey in the discussion section 3.5.

3.2 Early Development in the United Kingdom

At the start of the present period of research Goldschmidt and Goldschmidt (1973) produced an authoritative review of the then quite recent worldwide attempts at individualizing instruction in higher education. This review was described in chapter 2, where it was shown that the authors had located seven different systems for individualizing instruction in higher education; programmed instruction, computer assisted or managed instruction (CAI and CMI), Postlethwait's audio-tutorial and mini-course systems, schemes making use of applied contingency management, the Keller plan, individually prescribed instruction (I.P.I.), and finally the provision of a range of options in the instructional method used by individual students.
In the period leading up to this review, all seven approaches were not equally widespread, and the picture was constantly changing, with for example, programmed instruction declining and computer assisted instruction growing. However, during the present research, certain clearer trends have developed, both in the United Kingdom and worldwide. The Keller plan (and derivations of it), and C.A.I. have become far more established than any other system in the present context (see page 1/2). However, C.A.I. is beyond the scope of this thesis.

There are a number of reasons for this pattern. Programmed instruction continues to thrive in applications to which it is suited (Green (1976)), but these tend not to include whole university science courses. This is because such courses require high level objectives, and are expected to involve personal contact between staff and students, and retain student interest over a period of months. The range of adoption of the audio-tutorial system, I.P.I. and the provision of a range of instructional options is limited by a number of factors. Initially they require not only extra teacher time, but also extra material resources such as print and particularly audio-visual materials. The limiting factor is greater for I.P.I. and instructional method options, since both require a sufficient range of resources for individual students to choose their preferred learning method. Applied contingency management has not grown significantly because its practice has been framed in terms of the direct application of a controversial branch of psychology to teaching (Homme and Tosti (1971)), and as such it is mainly accessible to psychology teachers of a given persuasion.

The Keller plan's substantial impact is not easy to explain. It quickly became a well developed system, accessible to and workable by teachers of all kinds. It made demands mainly upon resources which go uncosted such as teacher time and secretarial help, rather than the "hardware" of educational technology (Rowntree (1974)). It was also flexible, being readily modified to fit many contexts in higher education. However, the influence of its basis in behavioural psychology is unclear; sub-section 3.4 2 shows that acceptance of this basis was never a reason for
teachers in the United Kingdom taking up the Keller plan. And experimentation has failed to conclusively link the success of the method to this theory of learning.

Up to this point it has not been necessary to restrict this review to the United Kingdom context. Indeed, with certain minor exceptions (e.g. Bori (1974)), all the developments described above took place in the Americas. The picture is one of innovation and experimentation, producing many systems for individualization during the years 1965 - 1970 as a result of widespread eagerness to apply learning theory to the practice of teaching an increasing body of students (Keller (1974)). This was followed by a period of development from 1970 - 1973 when certain methods began to dominate.

This was the setting for the transference of individual study courses, particularly the Keller plan (but also the audio-tutorial system) from the United States of America to the United Kingdom, and the remainder of chapter 3 describes the results of this transfer.

Early publication (Elton et al (1974)), and subsequent experience (Bridge and Elton (1977)) indicate that L.R.B. Elton first taught a Keller Plan course in the United Kingdom, at Surrey University in 1972 (Willoughby and Boud (1973)). In the years up to the survey, the Keller plan grew from this start, whilst at other institutions, different approaches to individualization were adopted, some of which were modifications of the Keller plan (e.g. Cohen and D'Inverno (1977)), and others were more closely related to the audio-tutorial method (e.g. Manwaring (1973)). This review cannot name or describe all of the resulting courses, but the way in which these innovations took root is important. This may be thought of as a three stage process.

The first stage 1972 - 1973 centred around the small number of individuals who pioneered the method in the United Kingdom. They attempted to fit the new methods into the existing context, and publicised their work at lectures and conferences. Much of this early work was done in collaboration with L.R.B. Elton, and was in undergraduate physics, mathematics, and chemistry at the University of Surrey (Elton et al (1974)).
The second stage started towards the end of 1975 when university teachers using or interested in individual study courses started to meet and disseminate their work in centralized groups. A "self-study group" formed within the Higher Education Learning Project in Physics, (Ogborn and Black (1973)), and was lead by L.R.B. Elton. Other less formalized group dissemination also took place by means of conferences on themes related to individualization and workshops for academic staff on the method, (Stace (1976)). At this stage the number of individual study courses grew steadily, although dissemination tended to be concentrated in certain academic subjects, resulting in the majority of courses being in the sciences, particularly physics. From 1973 - 1975 the variety of teaching methods emphasising individual study also increased as teachers fitted elements of the Keller plan into their corresponding variety of contexts. Although the Keller plan remained the most popular, two somewhat different approaches also emerged. The first was teacher-paced individual study (Cryer (1977)), which was different from Keller's self-paced study methods in that it put group work and regular class progress through the material in place of progressive mastery and freedom of pace. The second approach owed more to the audio-tutorial method, being based on a variety of resources, and often allowing freedoms of content and method as well as that of pace. Both will be described in section 3.4.

The third stage was just starting in 1975, and is still current. During this time experience, experimentation, and dissemination of information about individual study courses started to devolve, such that a number of centres of development grew up around one or more successful innovatory courses. Such centres are informal and go quite unacknowledged, so this assertion is hard to prove, yet the proliferation of innovative courses around the universities of Surrey and Glasgow (Hogg (1977)), and the polytechnics of the South Bank (Freemantle (1976)) and Manchester (Hollinshead (1977)) point strongly to the emergence of this stage. At this point the survey in 3.4 was undertaken.
3.3 The Methodology of the Survey

The survey had a number of aims. The most general was to gather together the existing experience of a family of innovations in higher education for the benefit of teaching staff either using, or considering use of these methods. It also had a summative purpose, in that it aimed to assess the extent to which the courses had achieved the different goals set by the teaching staff adopting them.

The survey was carried out during the first six months of 1975 for a number of reasons. Firstly, at that point it had become impossible to retain personal contact with the growing number of individual study courses. Secondly, by the time of the main circulation of the questionnaire in May 1975 all the respondents had at least two terms' experience with the method. Thirdly, a series of case studies had then been drafted by teachers in the HELP(P) group about their own experiences of running an individual study course, and these were used to indicate the range of issues which a wider teacher-survey might cover. Finally, as it was completed prior to the last year of the present research, the survey was timed to point out relevant research areas for the year.

Having formulated the aims and the timing of the survey, its scope and its method of reporting were decided upon. It was restricted to science and engineering teachers in higher education who ran whole courses which were to be learnt through individual study rather than traditional didactic lectures or seminars. The survey covered only the United Kingdom and Ireland because this was the focus of the research. And although previous surveys had been conducted in the United States of America (Hess and Sherman (1972)), the United Kingdom's distinct context had been omitted.

Reporting this survey promptly and accurately was considered vital if it was to satisfy the aims set for it. This was done in two stages. Firstly, all the respondents received a full report of the survey analysis in draft form in September 1975 (Bridge (1975)). They were thus able to use it in replanning their courses, but also they provided corrections of fact in the report, and feedback as to how the report could be made more useful to them. The report
was then extensively revised, and was published as promptly as possible in April 1976 (Bridge (1976)).

It was decided to perform the survey by questionnaire followed by informal interviews because this presented a relatively economical method of gathering and systemising data from respondents all over the United Kingdom. Furthermore the nature of the data to be gathered would be indicated by examining the case studies, and thus the somewhat inflexible questionnaire format was acceptable. Also it was considered unlikely that the target group would be prepared to spend time responding to a measure more demanding than a questionnaire.

To summarise, the survey procedure comprised a number of small steps:

(1) A group of initially six teachers in the HELP(P) project produced short personal case studies of their experience of using individual study courses. These teachers intended the case studies for publication (and this has subsequently taken place (Bridge and Elton (1977)), but I myself commissioned them.

(2) Having decided to perform such a survey by questionnaire and informal interviews, a draft questionnaire was produced and subjected to pilot study (Appendix I/1). Its coverage was derived from the case studies (above), and from an interim review of the research and literature on individual study produced by the present author in 1975 (Boud, Bridge and Willoughby (1975)). The resulting draft questionnaire was administered to three of the target group who taught at Surrey University, and were thus particularly accessible. Their difficulties in completing it and comments were noted.

(3) The questionnaire was revised on the basis of these comments, and it was then printed (Appendix I/1).

(4) A number of methods were used to locate teaching staff using individual study courses. The most productive method was by means of personal contacts, correspondence records, and mailing lists. Two other more systematic methods were also employed. The interim research and literature review was used to locate further suitable teachers, and the education groups of each of the learned scientific societies were circularised, and consulted on this matter.
(5) As a result, in May 1975 51 staff using individual study courses were located and sent the survey questionnaire. Some of these staff taught more than one individual study course, and they completed one questionnaire for each course. In all 59 questionnaires were issued, and of these 43 were returned for analysis.

(6) The analysis stage of this questionnaire was straightforward. Much of the data were factual or numerical such as the subject, level, or class size. The responses to open-ended questions were classified, and it was found particularly valuable to correlate these with the detailed method of individual study adopted by the teacher.

(7) During the analysis, responses were checked, and further questions posed in 14 cases by means of interviews over the telephone. The notes taken from these conversations were used in producing the first draft report.

(8) The first draft report was circulated to all respondents for comment and correction. It was also circulated to certain other university staff known to be interested in the results of the survey. On the basis of these replies the report was revised and finalized.

3.4 The Results of the Survey

The first question which the survey report tackled relates to the context within which the individual study courses grew up.

3.4 1. Who Teaches what by Individual Study?

Twenty-five of the individual study courses surveyed were in universities, nine were in polytechnics, and the remaining three were in colleges of education and of technology. Representation was remarkably even, with no particular emphasis on, say, technological or redbrick universities.

Eighteen courses were in the physical sciences, and of these eleven were in physics. Engineering was represented by nine courses and mathematics by five. Other subjects included mineralogy, computing and zoology. No similar courses in arts subjects were
located but there were five courses in the social sciences. Although all three undergraduate years were represented, the majority were introductory courses in the first year, as well as several pre-first year courses such as college 'A level' courses and the Scottish first year. There was also one M.Sc. course. Generally, the individual study course was only one of an average of five courses taken by students at any one time.

The amount of experience which teachers had of individual study courses varied widely. In 1975, some were still teaching their first individual study course whilst others had three or four years experience. The average was about two years.

The following table shows the range of class size. In many cases this had changed considerably from year to year.

<table>
<thead>
<tr>
<th>Class Size (Students)</th>
<th>Overall Percentage of Courses in Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 10</td>
<td>13%</td>
</tr>
<tr>
<td>11 - 20</td>
<td>27%</td>
</tr>
<tr>
<td>21 - 30</td>
<td>13%</td>
</tr>
<tr>
<td>31 - 40</td>
<td>11%</td>
</tr>
<tr>
<td>41 - 50</td>
<td>3%</td>
</tr>
<tr>
<td>51 - 60</td>
<td>6%</td>
</tr>
<tr>
<td>61 - 70</td>
<td>11%</td>
</tr>
<tr>
<td>71 - 80</td>
<td>8%</td>
</tr>
</tbody>
</table>

Table 3.1

Of the three courses not represented, two were in the range 141 - 150, and another was available optionally for about 800 students.

The second issue tackled by the survey was teachers' motivation for taking up individual study courses. In reporting it, I make use of extracts taken directly from questionnaire responses.

† The focus of this survey was courses in the sciences, therefore this should not be taken as firm evidence that such courses do not exist in Arts subjects.
3.4 2. Why did Teachers start to use Individual Study Courses?

Predictably, some teachers took up individual study courses mainly because they were dissatisfied with lectures and tutorials, while others were primarily attracted by their potential advantages rather than the disadvantages of traditional methods. These two motivations, although different were almost inseparable, and both had approximately equal support.

The dissatisfaction with traditional teaching occurred for a number of reasons, for example:

..... (I felt) "dissatisfied with the predominence of lecture based courses, and the widespread acceptance of 50% comprehension". ..... 

..... (I also) "believe that the pace at which students cover material is not uniform, and even the best lecture is" ..... 

Other teachers were dissatisfied with lectures in terms of the resulting examination marks; student, or staff comment. Five teachers said specifically that they were dissatisfied by their failure to deal with differing student prior knowledge when teaching by lectures.

Teachers were attracted to two potential advantages of individual study courses; firstly the possibility that they might improve (or, as some put it "make more efficient") the teaching process, and secondly the potential for increased independence of learning on the students' part. Under the first heading were ideas that individual ability and preknowledge differences could be dealt with in individual study courses; also that the course would benefit from regular feedback and progressive mastery tests. Also several wished to help weak students in particular, by means of individual help. Increasing students' independence of learning was sometimes seen as an aim in itself, but more often it resulted from a wish that students should learn how to arrange their studying effectively, or how to learn from scientific books. One respondent put it as follows:

..... "To place emphasis on individual study which would resemble more closely the kind of study expected of engineers in keeping abreast of new ideas".
There were many less universal aims. Individual study was often introduced to try to interest the students more in the subject, to present it in a more enjoyable fashion, and to motivate students to work hard on the particular course.

3.4 How were the Individual Study Courses Arranged?

This section is divided into three parts. The first covers the background situations into which the innovative courses were introduced. Here the data relate to individual study courses of all kinds, both Keller plan and individual study courses of a more diverse format. The second section concentrates solely on Keller plan courses, easily the largest group, and describes how the scheme was implemented. The third section gathers together descriptions of all the other approaches to individual study, and compares them with the Keller plan.

(a) The Background Situations

Thirty of the individual study courses had started as lecture courses and had been changed over by the teacher. Seven courses had been taught by individual study ever since the subject was introduced into the syllabus, or ever since the teacher himself had started teaching it. Of the remainder, three ran parallel to a continuing lecture course, and the others were previously presented by a series of seminars or tutorials.

The duration of these courses was equally divided between one term, two terms and three terms in length; on average 16 weeks. Almost all the courses were timetabled for between 1 and 4 hours per week, with 2 hours per week being the most common. Similarly, almost all the courses used 1 hour class periods, although in 9 cases, teachers had found that longer 1 - 2 hour periods were more convenient for courses run by the Keller plan.

(This contact time was used in a variant of ways. In Keller plan courses attendance at any one class period is optional, and staff/student contact on the basis of tutorial help or testing is 1:1. In other types of individual study courses attendance is often not optional, and the contact may be on an individual or group basis.)

The ways in which the individual study courses were staffed
varied enormously. One group of 14 teachers had no help in teaching their course at all. However, the largest group of teachers (17) were assisted by either one or two staff members or postgraduates. The remainder of the courses involved four or more teaching assistants. In the case of two large courses this help was simply carried over from the previous tutorial back-up to the lectures. However, in three courses the conversion to individual study had brought almost all of the department into active involvement in the course for the first time, when, for example, each staff member contributed one written unit or attended one class period per week.

In each case the preparation for the course had involved producing written or recorded units or modules. The majority of courses involved one unit per week, or just less, so that the average number of units was fourteen. One quarter of the courses also included optional units for use by the fastest students.

(b) Keller Plan Courses

Thirty teachers ran a course which corresponded to the fairly well defined format of the Keller plan, and thirteen teachers used courses which owed little or nothing to Keller's ideas (indeed some teachers ran both Keller plan courses and more general individual study courses).

It was important to ascertain why the majority of teachers chose to run individual study courses by Keller plan. The most common reason, (given by ten teachers), was that they had been in personal contact with someone who had previous experience with the Keller plan. Sabbatical visits, to and from the United States of America, and interuniversity projects were often stated in this context. Three individual teachers were encouraged to use the Keller plan by the success of the Open University course unit system, by the belief that it was based on sound psychology, and by the perceived need for progressive mastery of the material in certain courses.

In order to give a picture of these thirty courses, the teachers surveyed compared their own courses with the five features which Keller stated as basic to his plan. The results appear below:
(i) An important goal of the course is that the students should master each part before they move on. To do this they are tested before they are allowed to proceed and they must get almost all of the test right if they are to pass. Two thirds of those using the Keller plan noted that this was the case in their course. Some of these attempted to describe what they meant by mastery, for example by mentioning test pass marks of between 75% and 90%. The remaining ten or so teachers were not as rigorous as this. They allowed students to pass on to the next unit if they judged that the student would be able to cope with it, or if they felt that the student would learn the material better at some later stage.

(ii) The students are free to move through the course at their own pace according to their interest and ability in it. Just over one third of the teachers stated that this was true of their courses. The remainder brought to bear some pressure on the students, such as target dates or a recommended schedule. In addition, two teachers offered extra marks as an incentive to regular progress through the units. The others divided the units into several blocks, and all the students had to move from block to block at a given time.

(iii) The few optional lectures which are given are not essential parts of the course, and are regarded as a reward. Two Keller Plan courses deviated a long way from this. They ran parallel with a lecture course on the same subject for the same students, and the individual study course was regarded as a closely supervised tutorial system, or an optional addition to help the weaker students. Of the remaining 28 courses, about half contained no lectures at all: The most common reasons for this being the feeling that they would waste students' time, and the doubt on the part of the teacher that his or her students would regard such a lecture as a reward. In the remaining 14 courses, occasional lectures were given, although no one adopted Keller's idea that only the fast students should be allowed to attend the lectures.

(iv) All of the content of the course is in the form of printed materials such as written units and books. This was
universally true, however twelve teachers noted that, in addition to this they also include film, tapes, laboratory and computer exercises.

(v) Some of the test marking is done by undergraduate student tutors or "testers". Only two courses included this feature. In one case student testers were used exclusively, and in the other they were used as just part of the tutoring strength. The remainder of the teachers regarded this feature as unnecessary for their course, or impossible, or even quite undesirable. One course relied upon self-marking and group-marking, whilst another used computer marking. In the remainder, test marking was done by staff or post-graduate tutors.

Next, there is the question about the written units on these courses. In all but two cases a list of aims or objectives was included at the start of the unit, and in all but three a more general introduction to the unit was also usually included. Less common was a list of prior knowledge needed for the unit, being used by only half the teachers. On the statement of objectives, four teachers remarked that they were not convinced of their value to the students although one added that "they are essential for me!". However, a majority felt that objectives are useful to students. The course units then listed a procedure for learning the subject matter, such as by reading given sections in the textbook and solving problems. Many included notes on, for example, errors in the textbook and extensions or elaborations of derivations and arguments.

(c) Individual Study Courses not based on the Keller Plan

There were thirteen individual study courses which owed little or nothing to the Keller plan. The teachers of all but three of these courses had been aware of the possibility of using the Keller plan. Their main reason for not doing so was a feeling that their students could not cope with the autonomy given to them by the Keller plan. Two other reasons were the need for group work on the course, and the teaching of a subject in which progressive mastery was not appropriate.

There was a good deal more variation in the approaches to individual study taken by this group of teachers than by those
using the Keller plan. Overall, tests were used less, and non-print materials were used more. However there were two sub-groups taking different approaches to individual study.

The first, and more common approach made full use of written units very similar to those on Keller plan courses. There were compulsory tutorial group meetings each week, based solely upon the individual study unit. In most cases the students were expected to keep up with one unit per week (in other words the course was teacher-paced to a certain extent), and the class meeting resembled a tutorial supporting, say, a lecture course (occasionally with the teacher requiring the students to submit some form of homework). In two cases the pace was not firmly fixed, and the class periods reflected this, revolving around independent work and individual tutorial help. However, in the remainder of these courses the work rate was fixed by the teacher and they have become known as teacher-paced individual study courses.

The remaining, less common approach (used by three teachers on four courses) depended upon the use of a resources centre. In two cases this was open for much of the time. On these courses, students did much of their studying in the resources centre, using written units, tapes, slides, film and demonstrations. In these two cases there was a very wide choice of procedures for covering each topic in the syllabus. Generally speaking, if the centre was open for a long time, the teacher was not present all the time the room was in use. These courses were nearest to Postlethwait's audio-tutorial system, and were the least common form of individual study courses in the United Kingdom when the survey was conducted.

In both teacher-paced individual study, and the audio-tutorial courses, lectures were sometimes given. In six of the courses progress testing was formal, for example, homework was graded or class tests were held, whilst in the remainder, the teacher relied upon tutorial discussions to gauge individual and group progress.

The following three subsections describe findings as to the outcomes of the individual study courses surveyed. Subsection 3.4 described the assessment and evaluation techniques used on these
courses, and subsections 5 and 6 describe the conclusions and interpretations which teachers drew from them.

3.4 4. How did Teachers Judge the Success of these Courses?

The teachers measured the success or otherwise of their courses in two ways. Firstly, it was measured by means of student assessment procedures. On about half of the courses the students were assessed solely by a final examination at the end of the course, although it was often said that examiners’ meetings, for example, would take into consideration the number of units passed when judging borderline cases. In the remainder of the individual study courses, there was an examination, but passing course-work unit tests, or modules also counted towards a course-work mark which on average represented 35% of the total final mark. In two courses students were not assessed at all until two years later in the final examination, and in two others, students simply had to cover all of the units to pass the course.

The second measurement of success was through other forms of course evaluation. Nearly three quarters of the teachers surveyed had used questionnaires (most commonly), or interviews, progress records, or test results to monitor the student reaction to, and learning on the course. The results of assessment and evaluation appear in the next sections.

3.4 5. What were the Outcomes of these Courses?

The first question is whether the individual study courses achieved their aim of improving student learning (described in 3.4 2). The examination results were surveyed and they tell part of the story. Of the courses surveyed, there were no cases where the conversion to an individual study format had noticeably worsened the overall examination results (compared with a previous year, and/or a parallel control group of students covering the same material by traditional methods). However, in thirteen of the courses it was impossible to judge accurately the effect of the course upon the examination results, sometimes because the examination or the course itself had been changed to such an extent that the results were not compatible with any others, and sometimes
because, when the survey was carried out, the examination results had not yet emerged. However, a further thirteen teachers said that there had been a noticeable improvement in the examination results, and six teachers claimed a large and significant improvement had been made. In the remaining eleven courses, the teachers' reaction was that there had been little or no significant effect on the exam marks. Teachers in this latter group took a variety of views of this outcome, for example:

...... "but the results are in my hands anyway"
...... "I sometimes think you can teach any way and the results will stay the same"
...... "but I think I can improve them for next year"
and ...... "no change - alas".

However, in analysing the survey responses on this subject it became clear that the teachers placed at least as much weight upon their own subjective judgment of the course's success in terms of learning, as upon the examination results. Many felt that only this kind of judgment could take into account such factors as "good and bad years", the students general confidence in handling the subject matter, or changes in the assessment scheme. In a large majority of courses, the teacher's personal judgment was that overall, learning had improved.

There was however a group of nine courses which had failed to achieve one of their particular aims, that of helping and improving particularly weak students. However, in only two of these courses did the teacher feel that these students might have coped any better with the previous lecture system.

About half of the teachers surveyed felt that the most important outcome resulted from the ways in which students had to organise themselves and their study on the course. It was felt that this resulted in improved studying skills such as learning from books and making notes, and that the consequence was more purposeful, and harder individual work. There were many variations on this theme. The most common was that, through the formulation and statement of each unit's objectives, both staff and students had a precise idea of what was expected of them, and this helped to improve both the content and presentation of the course, as well as directing the students' individual study.
Ten of the teachers surveyed found that the most important outcome of their individual study course was the degree of staff/student contact in the class periods. Most of these teachers had used the Keller plan, and they found that the test sessions in particular furnished staff and students with a good deal of feedback about the course, and their learning on it.

The only other widespread outcome was a disappointment, felt by twelve of the teachers, that the individual study courses had failed to achieve certain ambitious goals such as "no real changes in their attitude to physics", and "no work beyond the mandatory objectives ..... students are highly skilled at playing the system and passing the exam".

3.4 6. What were the Reactions of Students, and other Staff to the Course?

All but five teachers reported that the student response to the introduction of an individual study course had been favourable. Of these five, two in particular set out with ambitious aims for students who were poorly motivated towards the subject. Two others noted that the students main complaint was that individual study tends to be time consuming, although this feeling was widespread even amongst the popular courses.

The students' reasons for liking individual study courses were the degree of staff/student contact involved, and the knowledge of exactly what was expected of them. Individual study was also seen as a "welcome relief" from a syllabus made up almost entirely of lectures.

The teachers were equally divided between those whose department's attitude towards individual study courses was neutral to favourable, and those whose attitude was indifferent or negative. In ten cases, initially sceptical departments were becoming more interested, and in nine cases, this is now resulting in other staff members also taking up individual study methods.

3.4 7. What did these Teachers Learn from the Experience of Running an Individual Study Course?

Two answers to this question were most common. The first related to time and commitment. Just over half the teachers had
found it vitally important to devote a generous amount of time to preparing and running their individual study course, (particularly for the first time), and they also felt that it was important that all those involved in jointly running the course should feel enthusiastic and committed to it. Some teachers learnt this the hard way when supporting staff withdrew, or when the fast students were held up because the units could not be produced quickly enough.

The other realisation was that, paradoxically, in the initial stages of running a course that requires students to work independently and on their own, it is important to make perfectly clear to the students what they are expected to do, and what they should learn. Again, a little over half the teachers felt that it is vitally important to have clear, straightforwardly written units which help students to do this, and in addition several suggested that, even in Keller plan courses which are supposedly self-paced, it is important to give students guidance as to a reasonable work-rate.

On the basis of these findings, many courses were reported to have undergone changes. The most common change has been a pruning of the course content, either by means of shorter or fewer units, or by making certain parts of the course optional. Also common has been an increase in the credit given for course-work such as passing a test or studying a particular exercise or module. This change, and the increasing use of "suggested" schedules has considerably reduced procrastination in Keller plan courses, a major initial problem.

No other changes were as widespread as these, and in different individual study courses, certain exactly opposite trends appeared. For example, in two cases teachers reported increasing the number of lectures given as a result of student pressure, whilst another two intend to decrease lectures or cut them out completely. Amongst the Keller plan courses there appears to be little trend away from testing, although the tests are valued more for the resulting staff/student contact and feedback, than for ensuring mastery (in whatever way it is defined). However, self-pacing is being increasingly restricted for those students whose natural pace is slow.
3.5 Discussion: The Emerging Trends

3.5 1. The Context

This survey has shown that individual study courses have been mainly used for introductory science and engineering undergraduates. There is a clear trend in context outwards, from courses in the physical sciences towards other pure and applied sciences; particularly mathematics and engineering. However, unlike the American experience with individual study, the context has virtually never extended into the humanities. The heavy emphasis on introductory pure and applied sciences is due to a number of factors. The first is that teaching based upon precise objectives (whether behavioural or not) appears to be more obvious and straightforward when the course is mainly concerned with the level of cognitive objectives commonly ascribed to basic courses in the sciences. Secondly, it was physical scientists rather than psychologists who were responsible for introducing the method into the United Kingdom, and European psychologists in general are less committed to the branches of psychology held to be the foundations of, for example, the Keller plan, than their American counterparts (Davies and Hartley (1972)). A third reason for the described trend is that in introductory science teaching there exist certain important needs which are matched by appropriate features of individual study courses. The most important needs are for a teaching method that can:

(i) adapt to differing student backgrounds
(ii) diagnose student weaknesses early on
(iii) place close personal contact before large group teaching in order to encourage the learning of material when it is first encountered.

Clearly such needs exist in many teaching situations, but in introductory science, the variety of student qualifications, and the structure of the subject matter play an especially large part.

Individual study courses have been a reasonably small-scale teaching innovation. Unlike the final year undergraduate research
project (Ogborn (1977)) or the undergraduate language laboratory, individual study courses have had little effect upon their home department, and have not significantly affected the staffing, budgeting, or timetabling of departments. However, recently certain established courses have broadened their impact. An example of this was the introductory mathematics course at South Bank Polytechnic which I evaluated in the first year of the research (Bridge and Laurillard (1975)). When this was converted to individual study format, the staffing was changed (to include the head of department), and the departmental tutoring system was geared to be able to act upon student progress data from the course. Later, an individual study course in mathematics at Southampton University had even more influence on the departments it served (Cohen and D'Inverno (1977)). Audio-tutorial style courses have also tended to result in greater departmental impact, in part due to their quite heavy demands upon resources (Manwaring (1977)).

3.5 2. The Procedures

A number of different methods for promoting individual study were brought to light in the survey. The main common factor was the use of previously prepared self-instructional materials as the means of conveying the content of the courses. In the majority of cases these took the form of printed units similar to those described by Green (1971), and now widely used elsewhere, for example, the Open University. However, in courses not based on the Keller plan such as the teacher-paced courses and audio-tutorial schemes, non-print resources like tape-slide sequences tended to play a larger part, often presenting students with a choice of learning method.

As well as learning materials, three other key variables emerge. The first is testing. Individual study courses were located which covered a range, from those which contained regular tests upon which progress through the material depended, to courses which were only tested once in the final examinations. However, the survey did not reveal any relationship between the testing policy of a course, and its outcomes such as examination successes,
although small Keller plan courses with their regular testing apparently resulted in particularly close staff-student contact.

The second key component is the pace at which individual students are presented with the material of the course, and this is closely linked with the testing policy. Four categories of pacing are suggested by the data:

(i) In the first, students move through the material of the course encouraged only by their personal drive, and general encouragement by the teacher (not backed up any direct rewards or sanctions). They gain nothing by completing the course-work quickly except the likelihood that they will do well in the final examination as a result.

(ii) In the second category, student work-rate is directly influenced by rewards provided by the teacher. These generally take the form of course-work credit obtained by passing each unit, (and are sometimes conditional on passing the units ahead of a given schedule).

(iii) The third category of pacing covers courses in which rewards for regular progress are replaced by methods for preventing slow progress, either by obliging students to move from one block of course units to the next at a given time or by arranging for essential parts of the course (such as experimental apparatus) to be available for only a limited period of time.

(iv) Finally, individual study courses may be teacher-paced in that, like more traditional approaches to university teaching, the material is presented to all the students at the same rate (whether they learn at this rate or not), and this rate is governed by the teacher.

I evaluated each of these categories of pacing during the early period of research. Introductory Chemistry (Freemantle (1976)) and Mathematics (Bridge and Laurillard (1975)) at South Bank Polytechnic, Classical Mechanics at Birkbeck College (Higgins (1977)), and Quantum Mechanics at Surrey and Strathclyde Universities (Boud et al (1975)) all employed the first category of pacing during the first year. Of these only the part-time students at Birkbeck College, and the students at Strathclyde maintained regular unit progress throughout the courses; in the remainder, poor coverage
of the units was very common. Both the Birkbeck and the Strathclyde students were unusually highly motivated; in the case of the part-time students (most of whom were studying in their spare time) this is essential for the whole degree programme. In the Strathclyde course a strong competitive group feeling developed, and the teacher there reported that this had also been noticeable on the previous lecture course. The second category of work-rate, where students are encouraged to progress regularly by course-work mark rewards, was evaluated extensively in the Particle Mechanics course at Surrey University. This substantially reduced procrastination, and is reported in chapter 4. The third category of work-rate was investigated in the Alternating Current course at Birmingham University (Black (1977)). This course contained theory and practical units, and a minimum pace was maintained by removing early experiments from the laboratory to make way for later ones, as well as by informal encouragement by the tutor. The result was regular student progress, with the slowest completing experimental units just before they were dismantled. Students accepted this imposed minimum pacing as a practical necessity. Finally, a teacher-paced course in Vibrations and Waves at Royal Holloway College was briefly evaluated (Kay (1977)). In this case, students received one printed individual study unit per week which they discussed but were not regularly tested upon. Thus procrastination in terms of not reaching later units was not possible. However, procrastination by failing to learn one unit prior to moving on to the next was certainly possible, although as in the lecture situation, this was hard to notice.

Lastly, the extent to which courses allowed for a choice of content was important. This was done either in order to allow students to follow their own special interests, or to provide alternative paths or goals for students of differing ability. Of the courses I evaluated early on, those in Alternating Current, Classical Mechanics, and Introductory Mathematics all contained at least one unit to be taken only by students with particular pre-knowledge difficulties. In the case of the Alternating Current course, students were routed according to pretest scores, but in the others, students were free to decide whether or not to make
use of the optional units. Conversely, the Particle Mechanics course presented students with a range of up to eighteen units from which they chose six or more according to their interests (see chapter 4).

3.5 3. Outcomes

The survey examined the outcomes of individual study courses which emerged in practice, and particularly those which corresponded to the teachers' aims for the courses. An important trend in the teachers' aims was observed, away from solely improving examination results, towards an attraction to intrinsic features of individual study courses such as their staff-student contact, their emphasis on the study of textbooks, or their potential for developing student independence. The data suggested that in the United Kingdom individual study courses have generally (but not universally) produced some improvement in examination results. This supports Stice's findings (Stice (1976)) rather than Kulik et al (1976) or Traveggia (1976), and thus adds weight to the argument in chapter 2 part II that the latter work may be flawed by differential commitment, and that a realistic evaluation of individual study course examination results should reflect quite frequent but not universal improvement.

However it was clear that well controlled educational experiments into learning as measured by traditional examination results have been lacking in the United Kingdom. Thus, although the teachers reported almost universally that they felt the students were learning more through individual study, the data bear out their feelings only tentatively.

There was no evidence to suggest that one particular procedure for arranging individual study courses resulted in better examination performance than others, and different levels and subjects taught also had little effect. However, both these conclusions remain tentative as long as experiments into the introduction of individual study courses continue to owe more to enthusiasm and pragmatism than to traditional, soundly framed psychometric research design. This is illustrated by the range of my tasks on the eight courses which I evaluated during the initial period of the research.
In only one case (P. Allen), was detailed analysis of examination performance jointly undertaken. Elsewhere, examination results, and the reactions of experienced staff assessing examination scripts were simply noted as satisfactory or otherwise.

Thus, if the aim of improved learning (in terms of the outcome of examinations), did not have a major impact, the outcomes of other, initially unintended aims must have justified the continuing growth of individual study courses. This was clearly so. Many of these were process aims rather than product aims (Cole (1972)), the most important being closer staff-student contact then traditional courses, and more autonomy for the student to direct his or her own studying. Such process aims have had a powerful effect. For although their results are both long-term, and hard to measure, it is clear from the detailed procedures of individual study courses that they are being tackled to a greater extent than is the case in traditional methods of university teaching.* It can thus be concluded that the continuing growth of individual study courses in the United Kingdom derives from the fact that they manifestly tackle process aims which are regarded as important for undergraduate science students, rather than from their proven success at achieving cognitive product aims.

Inevitably, certain aims were not matched by the outcomes of individual study courses. These included improving the performance of particularly weak students (although see Elton (1973)), and ambitious aims relating to students' attitudes towards the subject matter and studying in general. With hindsight, it is unrealistic to expect a single course to have much effect on the range of ambitious aims noted in 3.4.5, and many teachers acknowledged this.

The failure to improve the performance of particularly weak students is harder to explain. However, in individual study courses, 

* In the same way, experimental projects for final year science undergraduates manifestly tackle the process aim of giving students experience of the real tasks of an experimental scientist, and this being so, staff accept the considerable difficulty in assessing how far these long-term outcomes (products) are being achieved.
particularly those which allow some freedom of pace, the responsi-
bility for learning is placed firmly upon the student. It would be unsurprising if the weakest students found it more difficult to adjust to this than the more able students. Further, it has been widely reported that students on individual study courses believe that they require more effort than comparable traditional courses (Stice (1976)); and this may have a greater effect on the weaker students. Finally, and again mainly in the case of self-paced courses, the measures described for combatting procrastination result in a teaching method which is flexible and individualised primarily for the faster students, and thus they might be expected to gain an extra advantage on such courses.

3.5 Lessons

One of the main tasks of the survey was to assemble the range of lessons learnt by teachers in taking up and implementing individual study courses, for the benefit of other teachers considering a similar move. Two universal lessons emerged; that generous time and commitment is required on the part of participating staff, and that the learning materials should be clear and straightforward. Two further lessons come out of the majority of Keller plan courses; that the units should not contain as much material (or as much detail) as was "covered" in the previous lecture course, and that guidelines or incentives are required to ensure that weak students make progress through the units.

These lessons are born out by my early evaluation studies. Of the nine teachers I worked with, all but two needed considerably more preparation time for the individual study course than they would have expected for an equivalent lecture course. The two exceptions were the courses in Quantum Mechanics taught by B. Stace and P. Maas who took over course material produced by L.R.B. Elton. This actually produced a net saving in preparation time, and in the case of P. Maas, this was an important reason for his taking up the Keller plan.

Once prepared, none of the eight courses evaluated took less staff time to implement than would a traditional course; four courses required the same amount of time, and four required more. Generally,
large Keller plan courses (such as those taught by P. Allen and D. Laurillard) tended to require more extra staff time (compared with previous traditional courses) than small courses run by teacher-paced methods (such as that by S. Kay). This was true even where large Keller plan courses used undergraduate students as tutors (see the evaluation of Particle Mechanics in chapter 4).

The lesson as to the necessary clarity and straightforward nature of the materials was also illustrated in the eight early evaluations. In each case the material was modified from year to year, and in each case the goals were those of adding clarity or thoroughness rather than further extending or advancing the material. One example was the Introductory Mathematics course by D. Laurillard which started with four brief optional revision units in the first year of use (Bridge and Laurillard (1975)). Students had so much difficulty with these that in the second year it was necessary to extend them to full size units which students needing revision could work through and be fully tested on. The need for revision had been foreseen, but the needs for guidance in this had been underestimated. (A similar chain of events took place in the Engineering course at Lancaster cited in the first Monitorkit case study in chapter 5.)

Amongst the seven Keller plan courses evaluated in the early period, all presented less material than had previously been lectured. And after early difficulties with procrastination, all provided students with strong guidelines such as graphs or progress records in an attempt to promote reasonable rates of progress.

3.6 Conclusions and a Future Perspective

Individual study courses, particularly Keller plan courses are gaining a small foothold in introductory science in higher education in the United Kingdom. This survey described the way in which such courses have been put into practice, the context within which they have been fitted, and the lessons of experience gained by teachers. So much was straightforward. The major difficulty arose in discerning why individual study courses have
continued to grow, and deciding what achievements marked them out from other university teaching innovations in such a way that this growth should have taken place.

It has been impossible to relate the success of any particular course to the subject or level of students involved. Nor has any detailed method (such as the Keller plan, or the audio-tutorial system) emerged as more likely to produce successful results than any other. However, it has been possible to link different methods to different likely outcomes such as student-paced courses leading to advantages of staff-student contact, and problems with student procrastination, or teacher-paced courses resulting in greater coverage of the material and class cohesion, whilst sacrificing one-to-one contact between student and tutor.

Perhaps the main difficulty has been deciding what constitutes success in this area. The teachers' aims for taking up individual study courses (as described in section 3.4.2) ought to help in this, but there are two problems. Firstly, teachers' aims in running individual study courses have changed, and will continue to do so. Initially most teachers aimed to produce tangible cognitive improvements in, for example, final examination results. However, over time both the aims of individual teachers, and those of the entire group have changed such that long-term, affective, process aims are given highest priority. Thus it would be difficult to judge the success of individual study courses in a situation where teachers' priorities change so much. The second problem is equally serious. None of the teachers surveyed, irrespective of aim, had been able to arrange to test for the achievement of anticipated improvements in examination results under favourable experimental conditions. And furthermore although many had attempted to assess the achievement of affective aims, such as the development of student independence or staff-student contact, by means of course evaluation, the nature of such aims has made the interpretation of the results very difficult indeed. Therefore, it has been suggested in this chapter that the reason for the apparent success of individual study courses (and their consequent expansion)
had not been a result of measured achievement of desirable product aims. Rather it has been due to the fact that the detailed learning pattern required of students in individual study courses corresponds more closely to certain important process aims in undergraduate science teaching than do the processes in other more traditional forms of university teaching.

However, there were other achievements located by the survey. It may be argued that the key effect of individual study courses was upon teachers rather than students. For teaching by this method has clearly prompted in staff a far greater awareness of the processes of teaching and learning than is the rule in traditional courses. The need to state clear objectives, to commit one's teaching to paper, to deal with student difficulties on a regular 1:1 basis, and to carefully monitor student progress through the course material all force the teacher to reconsider how, and why each part of the subject matter should be taught. For these reasons, it was predictable that the survey should locate a number of science teachers in higher education who were devoting the generous amount of time and energy to teaching which one might otherwise expect to be taken up by academic research. For such teachers, the existence of tried innovations such as the Keller plan allows a reasonably secure first step away from traditional teaching. And this has often been followed by further changes and development in the light of experience.

This survey was both a piece of research in its own right, and also a source of projection about further research into individual study courses. Certain important factors arise from this. The most important is that, up to the time of the survey, the emphasis of research and development into individual study courses has rested upon the formative development of viable teaching processes. Whilst this is probably appropriate for the early stages of any teaching innovation, the survey underlines the idea that future research might profitably adopt a summative role, and, while accepting that many teachers continue to use individual study courses because they satisfy their own process aims, undertake to discover the products or the achievements
of such courses in terms of cognitive and affective changes in the students.

The second point which the survey makes is that a number of science teachers in higher education are both able and willing to undertake pragmatic educational research and development. Whether or not they are encouraged to do so, they modify and change teaching methods to fit their own needs, and evaluate the outcomes quite critically. Indeed, the diversity of experiences reported in the survey suggests very strongly that those wishing to promote innovation in university teaching will both encounter resistance, and fail to tap resources of teacher skill and enthusiasm, if they do so via unflexible pre-prepared schemes which allow little room for adaptation.

Therefore, in looking to future research and development of individual study courses, it is important to harness teacher research in at least one of the two areas of research which now appear necessary. The range of approaches to individual study is growing, and it is important that variations, such as the use of computer management, or courses which combine teacher and student pacing, or courses based on a learning centre, be evaluated as to the effects of these modifications upon the core notion of individual study. In this teachers themselves may play a large part. The second research need is for more overall evaluation of the family of innovations which have been developed. The foundations for the former teacher research have already been laid in the development of an evaluation resource package for teachers described in chapter 5. And chapter 6 addresses certain overall questions about the individual study courses which have been developed.

Finally, the survey prompts certain projections as to the future of individual study courses. As stated before, the most important trend seems likely to be diversification of method, such that the basic idea that students should be made more active and autonomous through individual study results in a range of modest innovations such as teacher-paced courses, more ambitious student-paced courses, and occasionally major changes on an institutional level such as those required in the audio-tutorial
style courses surveyed. And it is reasonable to suggest that, amongst these, small teacher-paced courses and courses allowing restricted freedom of pace, being the most economical in many respects, will predominate. However, it is important to accept the likelihood that by growing significantly, this innovation like others before it will be subject to criticism and adverse scrutiny (Young (1974)). In facing this, United Kingdom courses may have one important advantage over their American counterparts. In the United Kingdom there is a wide variety of such courses all founded upon a very general, but widely acceptable ideal, that active individual study is one important method for teaching and learning science. In the United States of America the picture is different. Most courses deviate very little from a small number of methods; the Keller plan and audio-tutorial scheme in particular (see chapter 2 part II). And for Keller at least, the justification for the innovation is in its application of theories of learning which however, are not universally held (Keller (1974)). Therefore, individual study courses in the United Kingdom may present a more diffuse and less controversial target than those in the United States of America.
CHAPTER 4: CASE STUDY OF THE DEVELOPMENT AND EVALUATION OF AN INDIVIDUAL STUDY COURSE

4.1 Introduction
4.2 Particle Mechanics 1973. The First Trial
4.3 1974 A Re-run in the Light of Experience
4.4 The 1975 Course: Little Changed for the Third Year
4.5 1975 - 1976 The Final Year of the Case Study
4.6 Evaluating Particle Mechanics
4.7 Conclusions
4.1 Introduction

The chapters up to this point have shown how individual study courses have progressed from being rare, unusual and carefully scrutinised experiments to their present position as a significant innovation in university science teaching. As the courses became more established, the extent to which they were experimented upon and evaluated from outside reduced markedly.

One course at least was an exception; first year Particle Mechanics for students of Electrical Engineering and Physical Science at the University of Surrey. This course has undergone regular revision and change during its four years in individual study form. It is also the only course that I have studied over the entire period of the research. Hence a record of the course and its evaluation forms the major case study in this thesis.

The case study is presented in two parts. The first four sections 4.2 to 4.5 are effectively evaluation reports describing the history of the course itself. The second part in 4.6 describes the course's evaluation.

4.1 1. Difficulties with the Data

The problem highlighted in the introductory chapter 1 appears throughout this chapter. This is the difficulty of backing up assertions, confirming observations, and generally allowing the reader to trace back from conclusions, through methods of analysis, to source data. The problem is compounded by the nature of the evaluation which owed far more to field research (chapter 2) than psychometric approaches. The resulting data are not concise, being only rarely numerical, and it is particularly difficult to present them in an accessible form.

The problem has been faced in a variety of contexts, including that of the present thesis. In fact I adopted different solutions when reporting the evaluation to the teacher himself, to the other staff on the course, the other members of the HELP project, and other wider audiences (contrast Boud and Bridge (1974), Bridge (1974), and Bridge (1975)).

For the purpose of this thesis, I have divided the source
materials into three categories. Firstly certain data are both vital and concise, and these are included in the text of the chapter. Secondly, certain materials are required to illustrate key points made in the text, and these appear in the appendices. Thirdly, data and data summaries not required for central themes are to be found in supporting documents. A weighty fourth category exists also, and this will not be referenced. This includes the raw data in the form of individual questionnaire responses, record sheets, interview tapes, and examination scripts, which have not been destroyed, and could be made available.

4.1 2. The Context of the Case Study

The context of an evaluation such as this is most important, for it prescribes what can be aimed for, what can be done, what range of alternatives can be considered, and what constitutes a successful outcome to a very large extent. In this section the context is simply presented. Section 4.6 assesses the impact of the context upon the evaluation.

A key context variable is the people and the relationships involved. I acted as evaluator on the course during the latter three years of the course's existence, but my position was complicated by being also a research student in the department headed by the course teacher L.R.B. Elton, and being supervised by him towards a higher degree. Furthermore, I was a member of the HELP project, working in a team devoted to the development of individual study courses, and to that extent the range of alternatives open to me was focussed. Finally I had only a general grasp of the subject matter being taught, whilst the teacher is an authority on it. He was strongly committed to the innovative teaching method, and was also keen both to conduct educational experimentation and to have the results evaluated.

Also important is the fact that the first year of the evaluation was conducted by another research student, L. Willoughby. The methods and outcomes of her evaluation produced expectations and limitations which had to be adapted to.

Particle Mechanics is a service course provided by the Institute for Educational Technology, left over from a larger interdisciplinary
service course programme which was gradually abandoned. Until 1975 - 1976 contact between the course and the departments it served were loose and sometimes uneasy, such that for example, tutors in the home departments doubted their responsibility for helping students with difficulties in Particle Mechanics.

4.1 3. The Background to the Course and the Aims of the Innovation

Particle Mechanics is a two term course for first year students of Electrical Engineering and Physical Science. It has been taught by L.R.B. Elton for the past eight years, except during 1971 and 1972. Until 1970 it was a lecture course attended by students from a number of departments. After 1972 it was decided to teach the course by means of the Keller plan. The decision was primarily that of L.R.B. Elton and he continued to be responsible for the course which catered for fifty to seventy Electrical Engineering students and seven to fourteen Physical Science students during the period 1973 - 1976.

The first three courses took place in the second and third terms of the first year programme, but the fourth course occupied the first and second terms. This accounts for the timing of the four courses as: 1973, 1974, early 1975, and late 1975 - 1976.

At the time of the transition from lecture to Keller plan format the teacher had the following aims in mind (Willoughby and Boud (1973)):

(i) Firstly, it was intended that students should develop confidence in the subject by studying individually, and learn how to organise their study time.

(ii) It was anticipated that Keller's concepts of freedom of pace and progressive mastery testing would result in the students learning the basic material more thoroughly.

(iii) Staff/student contact time should be used in a purposeful way. The unit tests were seen as an opportunity for informed and focussed discussions between staff and students.

(iv) Keller plan courses had been used with some success on two smaller courses the previous year, and the teacher was anxious to experiment with its use on a larger course, and to tackle the
doubt that Keller plan courses were impracticable for large classes.

Of these aims, the two most commonly repeated after the initial trials were that the results (usually the examination results) of earlier lecture courses were unsatisfactory and thus change was required; and that it was intended to try out the Keller plan in a large class.

4.2 Particle Mechanics 1973. The First Trial

4.2 1. Plans and Procedures†

Having decided to run Particle Mechanics as a student-paced individual study course, it was first necessary for the teacher to set down on paper the material which he previously lectured. The first step was to locate one or two books upon which the study guide units might be based, and which could be referred to throughout the course. The book chosen was one written by the teacher himself (Elton (1971)).

Twelve separate units were produced, each containing an amount of material similar to that previously covered in one or two weeks' lectures. Two forms of a test were produced, in order that students who failed one would not have to retake exactly the same test later. A sequence of three optional lectures were prepared on the subject "time". In view of the considerable amount of paperwork required to run such a course for seventy students, it was decided to pay a post-graduate student (L. Willoughby) to act as administrator for the course, responsible for handling the study guides, the tests, and other day-to-day matters. She also acted as evaluator on the course, and some of the details of the 1973 course which follow were first reported by her.

The course started in spring 1973 with an introductory lecture where the students heard details of the course and advice as to how they should work through it. The one hour testing and

† It should be noted that I played no part in the evaluation of the 1973 course. These brief records of the 1973 trial in 4.2 have been gathered from documentary material produced at the time.
tutoring periods held twice a week in a classroom seating only thirty or forty students were very crowded during the first few weeks of the course. Long queues of students waiting to have their tests marked soon formed each period since, at first, about half the class turned up to have the early units marked by a tutor force of only four post-graduates and one staff member.

Once the students had passed one test they were given the next study guide. Test sessions involved plenty of discussion on the subject and the solutions in question. There was no penalty for failing a test; an equivalent form simply had to be retaken. However, students were told to aim to pass all the units by the end of the course, when a traditional examination (on which the grades would be based), was to be held. The final component of the course was a "term paper", an essay on a subject related to particle mechanics which each student had to write during the Easter vacation.

4.2 The Outcomes of the 1973 Course

In summer 1973 L. Willoughby reported on the evaluation of the first course (Willoughby (1973)), and I started to work in partnership with her. This ran into 1974 when I took over the evaluation of the course and she continued as course manager.

The evaluation of the course had considered a number of areas. Two were of particular interest. Firstly, student progress through the twelve units was examined, (see figure 4.1). As can be seen, few students had completed all the twelve units by the end of the course, and the majority had got less than half way. This lack of unit progress was regarded as the principal problem with the course by all concerned, and informal interviews by L. Willoughby with the students suggested that this was due to the students feeling a lack of routine, or any compulsion to work.

Secondly, attendance levels at the class periods were monitored, and although this varied from week to week, there was a clear decline in the number of students attending class as the course progressed (see figure 4.2). The individual student progress records kept by the evaluator showed that this was due to the students who had not progressed very far virtually ceasing to attend
Figure 4.1 Students' Unit Progress

1973

1974

1975 (Spring)

1975-6

Total Number of Units Completed in Year
Unit passing deadline

Deadline fixed in error and changed

(Not all students realised this)

End of First Term

Week Number
at all, whilst the faster students continued taking tests at a steady rate. This high attrition rate was attributed to the queues and the general pressure on the small number of tutors available which resulted, particularly for the weak students, in a lot of time wasting followed by a short, hurried test or tutorial session when they attended the class.

The students' examination performance was investigated. Although they were certainly no worse than the results from previous lecture courses, and the teacher felt that there had been some improvement in learning, the numerical results did not reflect this in terms of strongly significant improvements. However, by this time the teacher's aims for the innovation and evaluation had changed (chapter 2 quoted him as saying "we often feel it in our bones that there are beneficial changes of a less tangible kind, if only we could pin them down ..... (by means of) new evaluation methods" (Elton (1973)). So other outcomes were seen as justifying the continuation of the Keller plan. As might be expected, the students who did well on other conventional courses also tended to be the ones who did well in the Particle Mechanics examination. A short questionnaire was issued to the students at the end of the course by L. Willoughby (A†). For them the greatest success of the course was the level of contact between the staff and the students. Unfortunately it was true that this advantage only emerged when many students had been put off attending the course by the queues and crowding.

Two other student reactions were common. Firstly, it was widely felt that the course was more time consuming than traditional courses (a theme which was first seen in chapter 2, and will be addressed specifically in chapter 6). Secondly, the idea that students should be confident that they had "mastered" a unit when they passed its test was not always realised. Although the students valued the feedback which the tests gave them, about half felt that the tests were not thorough enough.

† The evaluation procedures used in this study have been coded to help trace the source of the results. Each separate method has a capital letter code. Table A.1 which appears overleaf contains a list of these methods with details of their scope, and a key for referring back to this part of the chapter on evaluation results.
<table>
<thead>
<tr>
<th>Year</th>
<th>Code</th>
<th>Evaluation Format</th>
<th>Details</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973</td>
<td>A</td>
<td>General Course Questionnaire</td>
<td>Questionnaire issued in last week of course. Response rate 90%. Questions derived from involvement in course.</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Informal Student Discussion</td>
<td>Held throughout course. Short (less than 20 minutes), individual and group. 33% of students covered.</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Class Observation</td>
<td>Informal, but covering all the class periods.</td>
<td>-</td>
</tr>
<tr>
<td>1974</td>
<td>D</td>
<td>First General Course Questionnaire</td>
<td>Issued at end of first term. Response rate 82%. Questions derived from class observation.</td>
<td>Appendix I/2</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Second General Course Questionnaire</td>
<td>Issued at end of second term. Response rate 58%. Shorter than D, prime aim to cross check interview data (F)</td>
<td>A Supporting Document (S.D.)</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Formal Student and Staff Interviews</td>
<td>Held throughout course. Long (45 minutes), individual, 40% of students covered. Interview schedule developed tape recording used.</td>
<td>Interview Schedule a S.D.</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Class Observation</td>
<td>Formalised, with a standard format. Recorded in diary. Used throughout course.</td>
<td>Standard diary format a S.D.</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>Student Progress Data</td>
<td>Class unit progress monitored at intervals. Also selection of optional units recorded.</td>
<td>See figure 4.1</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>Evaluation of Optional Lectures</td>
<td>An individual project within 1974 evaluation programme. Based on standard format observation scheme and teacher/evaluator discussions. Project developed over the three optional lectures.</td>
<td>For observation scheme see Casper (1975)</td>
</tr>
<tr>
<td>1975</td>
<td>K</td>
<td>Class Observation</td>
<td>As G</td>
<td>As G</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>General Course Questionnaire</td>
<td>Issued at end of second term. Response rate 72%. Questions aimed at illuminating causes and extent of dissent (see 4.4)</td>
<td>S.D.</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Student Progress Data</td>
<td>As H</td>
<td>As H</td>
</tr>
<tr>
<td>Year</td>
<td>Code</td>
<td>Evaluation Format</td>
<td>Details</td>
<td>Reference</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>1975</td>
<td>N</td>
<td>Recording Test/ Tutorials</td>
<td>36 test/tutorial sessions held by staff, postgraduate, and undergraduate tutors were tape-recorded; at a range of points in the course, and on a range of unit topics.</td>
<td>See Chap. 6 Also sample in Appendix I/4</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Student Diary Books</td>
<td>Issued to 30% of the students on the course, these showed records of students private study workload for the period of 7 days.</td>
<td>Chap. 6 Also sample Diary in Appendix I/3</td>
</tr>
<tr>
<td>1975-6</td>
<td>Q</td>
<td>General Course Questionnaire</td>
<td>Issued at end of first term. Response rate 71%. Aimed at assessing the extent to which previous problems overcome, and the impact of further new features of the course.</td>
<td>S.D.</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Class Observation</td>
<td>As G. Standard format further developed.</td>
<td>As G</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>Student Progress Data</td>
<td>As H. Primary focus the effect of removing coursework credit from deadlines.</td>
<td>As H</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>Pilot Questionnaire and Interview Study</td>
<td>Questionnaire on study methods issued to 60% of physical science students who were then interviewed on basis of responses, at end of first term.</td>
<td>S.D. Also See Bridge &amp; Hodgson (1976)</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>Formal Student Interviews</td>
<td>Held at start of second term. Long (30-45 minutes), individual, 48% of all students covered. Based on interview schedule developed from T, and on Q.</td>
<td></td>
</tr>
</tbody>
</table>
Overall, the evaluation showed that the 1973 course was neither a great success nor a failure and it was felt that the problems relating to student routine, and tutor:student ratio could be overcome.

4.3 1974 A Rerun in the Light of Experience

4.3 1. Plans and Procedures

From summer 1973 when L. Willoughby handed over evaluation responsibility to me, and submitted the evaluation report on the 1973 course to its teacher, plans were made to change and develop the course. There were two clearly defined problems to be tackled by changes in the format of the course. These were the need for extra tutors on the course, and the need to provide the slower students with some incentive and routine for at least a minimum amount of regular work.

The main change which was made was the introduction of a course work mark-scheme which resulted in students' final grade being made up of both course-work and examination results. It was hoped that this would provide more incentive and routine for the weaker students in the following ways. Students received a unit progress table along with other course documentation at the beginning of the course. This table identified dates fixed at fortnightly intervals throughout the course, and students received one course-work mark each time they passed a given unit on or before these deadlines. It was decided to make eight units the expected minimum for the course, rather than requiring all the students to cover all twelve, and so eight corresponding deadline dates were laid down. The total possible number of marks for the whole course including the examination was 128, so passing units in this way provided rather nominal credit.

The mark scheme was also used to increase the number of tutors on the course, for it was decided to ask the fastest students to test their colleagues on units they themselves had already passed, again in return for one course-work mark. This had a
variety of aims; certainly one was the wish to improve the tutor-student ratio, but another was enthusiasm on the part of the teacher to research and innovate in this approach to teaching.

There was one further very major change made to the course which did not spring directly from problems isolated the previous year. Instead of covering the course units in a linear sequence, a branching scheme was devised so that, after covering a core of two units, students had some choice in which units they covered. This was intended to help the students build up a course suitable and relevant to their own needs, particularly in view of the two groups of students (Engineers and Physical Scientists) on the course.

The final change was to reduce the length of some of the later units by splitting off the more advanced part. This shortening was done partly to reduce student workload towards the end of the course, and partly because the later units had often occupied the students for up to three weeks, and this was felt to result in an unacceptable period of time between tutor contacts.

With these changes the course was rerun in 1974. The students were introduced to the course in a similar way to 1973, but they were presented with all the course units bound together at the start of the term. This was done so that they would be able to make an informed choice amongst the alternative units later in the course.

Class attendance during the first few weeks was again high, but a determined effort had been made to cater for this. A larger room was used, and the number of staff and post-graduate tutors on the course was built up by using temporary and visiting staff, undergraduate testers were also recruited from amongst the fast students. Between two and four students testers assisted the four to six staff and post-graduate tutors usually present during the class periods.

Organising student testers was done by the course administrator, L. Willoughby, and she found that the branching system produced problems when it came to choosing them (Willoughby (1974)). It was also felt that undergraduate testers should be prepared for
their job, and so interested students were invited to a short talk
about testing held at the beginning of term. In class, student
testers were given the unit test solution sheets, and they worked
at separate desks at the front of the class in the same way as
the staff and post-graduate tutors.

A final change to the course for 1974 was the inclusion of
optional computer exercises for some of the units. Like the labora-
tory demonstrations used throughout the course, these were intended
to be largely self-instructional. Students received two course-
work marks for each of the four computer exercises they undertook.

4.3 2. 1974 Outcomes

The modified and revised course run in 1974 was regarded as
a significant improvement over the 1973 course. Records kept in
the evaluation (figures 4.1 and 4.2) show that the key problems
from the previous year had been tackled, in that class attendance
and unit progress did not deteriorate throughout the year. The
evaluation also suggested that the course was popular with most of
the students who found the freedom of content resulting from the
branching scheme attractive, indeed, more so than the freedom of
pace (E and F).

As can be seen from figures 4.1 and 4.2, in 1974 a much
higher proportion of the students on the course made a genuine
attempt at it, and the majority completed at least a minimum of
eight units. The attendance figures were also improved, with no
marked drop off over the two terms. The effect of deadlines can
be seen in figure 4.2 which shows that attendance peaks usually
occurred at deadlines. In their responses to an end of course
questionnaire (E), the students said that this improvement was due
to having a routine, and knowing what was expected of them, rather
than through a spirit of hunting for extra marks.

The experiment with student testers started uneasily, but the
results improved throughout the course. From the testers' point
of view the exercise had been valuable. Many remarked in the first
general course questionnaire (D) how much they had learnt, realised,
and revised through testing. However, they did feel that they
should have been properly prepared for the job of testing. Many had not attended the talk on testing given at the start of term, and those who had soon forgot the details. However, by the end of term, the regular testers felt that they had learnt by their mistakes. These initial problems were reflected in the feelings of the students who received help from undergraduate testers. At first their attitude was somewhat negative, feeling that the testers were not tough enough, and that they stuck too rigidly to the answers on the solution sheets (D). However, by the second term this feeling had changed. Many students then preferred the student testers, finding them more approachable, with a better understanding of their problems. This was checked and confirmed in the second short questionnaire issued at the end of the course (E).

However, the problem of student workload was not improved at all, particularly in that many of the weak students remained working on the course in 1974, rather than dropping out. Furthermore, the marking scheme offered extra marks for laboratory and computing exercises, and also for advanced units, and some weak students who would have been better employed working on the core of units tended to spend much time on "extras" (G).

A second problem which was not solved in the 1974 course was that of student confidence in the unit material after they had passed its test. This did not seem to vary with the tutor or student testers concerned, and feeling again was that the tests were not thorough enough (G and F).

Both in 1973 and 1974 the optional lectures had a predictable effect. They were regarded as interesting but not very useful by the students (questionnaire F and the separate evaluation of optional lectures J), and in 1973 and 1974 the attendance gradually dropped. Similarly the computer exercises were regarded by most students as interesting but not always directly relevant.

Overall, the staff were happier with the 1974 course, and it was decided to concentrate upon consolidation in the following year. The examination results had again been considered. Once again they were a slight improvement over the results obtained in the lecture
course, but the significance of this was not checked, and the teacher's own impression that the students were learning more by the Keller plan was regarded as an equally important assessment. This was particularly true since there had been considerable changes in the material which the students covered due to the branching scheme.

4.4 The 1975 Course: Little Changed for the Third Year

4.4 1. Modifications and Changes

The course did not appear to need any major changes on the basis of the 1974 evaluation. Of the three problems left over from 1974, that of the undergraduate testers' initial difficulties with tutoring was tackled by producing an instruction sheet for potential testers to study prior to testing. Much of the advice it contained stemmed from experience with testers the previous year. A routine was also established whereby the teacher checked the marking done by testers. In an attempt to reduce the workload, and concentrate the students (particularly the weaker students) on the core of the course, the mark scheme was altered to reduce the credit for optional extras such as computer exercises. Finally, the unit tests were standardised, so that each contained four questions of a given format. This meant that clear criteria for test passing could be set (three questions perfect plus a good attempt at the last one). It was anticipated that this would both boost students' confidence in their learning of any given unit, and make the student testers' job more straightforward.

Other modifications were also implemented: the minimum core of units, and the corresponding rewarded deadlines were increased from eight to nine in the hope of increasing the students' coverage, and it was decided to hold the final examination twice. This latter step was to encourage students to work quickly and finish the minimum number of units early. This qualified them to take an early examination in May, which would reduce their load during the normal examination period in June. Finally, the students' range of choice
in selecting areas of study from the branching scheme was reduced. Guidelines were laid down showing the most suitable units for the two groups of students to take, in the light of advice from the home departments.

4.4 2. Difficulty in Practice

By 1975 the course tutoring had settled at two staff, two post-graduates, and two visiting staff members who were not always present. Undergraduate testers were also used again. Finally, it was decided that the evaluation of the 1975 course should be less intense than in 1974, and therefore I took on both the roles of evaluator and administrator.

Again, students were given details of the course such as its assessment and the organisation of tests and test sessions at an introductory lecture. The Keller plan and the subject itself were well received and popular for the first few weeks, as in previous years (class observation records: K). Towards the end of the first term, its popularity dropped as students found the course hard work, and the weaker students became aware of their relatively poor progress. For a significant group of students, this disenchantment increased over the second term, and the course ended with discussions at department level as to what had gone wrong. This pattern was not typical, students on this course in the past had lost their mid-course scepticism towards the end when they found that the progressive testing and assessment made examination taking less critical (evaluation procedures G and K). The courses, processes, and outcomes of this dissent will be discussed below, but having set the scene, an evaluation of the less controversial facets of the course is first presented.

As before, the students received all the course units in a printed book prior to the start of the course. The revised branching scheme left students a fair measure of freedom over their sequence of study, and naturally certain units were popular because they were reputed to be easy or interesting whilst others were avoided as being too difficult (Bridge (1974)).

Students again had the option of doing laboratory work or computing exercises. As in 1974 they tended to find computing the
more enjoyable, but the laboratory exercises were, they felt, more relevant to the course (L). Fewer course-work marks were available for passing computing exercises and this did dissuade weaker students from spending their time on computing rather than the more important core of reading units. However, in total fifty-seven computer exercises were undertaken by students on the course.

Two further teaching components were the term paper and the optional lectures. In 1975 the term paper caused some difficulties. Firstly, many students spent a lot of time on it (as much as a whole week of the Easter vacation), and of these, several were dissatisfied with the marks they received for this (K and L). Secondly, it was believed that certain students had "got away" with copying their essay directly from a textbook, and been awarded good marks for it.

There were again four optional lectures on the subject "time". Attendance started high but declined more markedly than before, such that 60% of the class attended the first lecture, but only 15% attended the last. Much more than in previous years, it was felt that the lectures should be related to the course directly, a common reaction being that optional extras were of dubious value in a service course (L).

During test periods in 1975 it was noticeable that very few students obtained more general tutorial help from the staff or post-graduates (recorded in the class observation records: K). This was not due to crowding or queueing since the move to a larger room, and the increasing use of undergraduates as testers meant that queueing and crowding were less than ever before. Students felt that they could not get help from the tutors as they were not amenable to general tutoring (as opposed to marking tests), and they pointed out that it was hard for them to admit to having problems and ask for help.

The sequence of deadlines again provided course-work credit for passing units. In 1975 the minimum had been increased from eight to nine units, but the average progress and attendance worsened slightly, although it was still better than 1973 (see figures 4.1 and 4.2). The final two schedule deadlines were so late as to be virtually ignored by many students.
There was some indication from the general course questionnaire (L) that students were more confident in their knowledge of the units once they had passed the standard format tests than had been the case previously. Also, this made test marking easier, particularly for undergraduate testers. The previous year when undergraduate testers were used for the first time, the initial student reaction was somewhat negative, and for 1975 a sheet of instructions and advice was prepared and used by each tester. As a result, the initial use of testers went very smoothly and they became an accepted part of the system.

The students' choice in the subject and order of units was limited over the 1974 scheme by the requirements of the examination, by guidelines set down by the home departments, and by changing the status of certain units to being unexamined options. The result was that the paths through the units chosen by most students were very similar, and corresponded to the guidelines set down in the introduction to the course. However, this restriction did not prevent the element of choice reappearing as a primary advantage of the 1975 course in the general course questionnaire (L). There were two reasons for this. Firstly, although slow students who only wanted to cover the bare minimum of the course had little choice as to which units to study, they did have a good deal of say in when they studied them, and they could move on from a unit which they found hard or uninteresting and return to it later. Secondly, a substantial degree of choice remained for the faster students, and once they had completed the recommended minimum number of units, they were able to take further units, and do laboratory or computing work.

Turning now to the student dissent, and complaints about the course towards the end, it is clear that most of the small and well intentioned changes made for 1975 had undesirable yet unforeseen outcomes. The changes in the assessment scheme had a key effect in this respect.

There were nine rather than eight deadlines, each earning one course-work mark. This, and the fact that computer exercises and optional units were awarded only one mark each, whereas in
1974 they had earned two resulted in a greater proportion of the course-work marks depending on test passing. The second change to the assessment scheme was holding the final examination twice, to encourage students (particularly the more able) to complete the course quickly.

In almost all respects, these changes which were intended to improve the overall performance had the opposite effect. There were a few fast students who appreciated the facility of an early examination, and in particular the resulting possibility of re-taking the examination to improve their marks (the general course questionnaire (L) isolates seven such students). However, it was almost universally felt that the early examination was unfair. A common explanatory comment (extracted from class observation records K) was:

"surely it should be the weak students, not the best ones who are allowed to have two tries of the examination".

Similarly the increased minimum core of units, and the greater proportion of course-work marks dependent on test passing within a given schedule had few good effects. The students soon learnt from second year students that the minimum had been increased from eight to nine for 1975, and they regarded this as unfair extra work, for example:

"it gives more marks to the best people, but just makes more work for the not so good". (Extract from K).

And because unit passing marks counted more heavily towards the final grade, this also provided an incentive for cheating.

As evaluator, I monitored the outcome of these factors; the following paragraphs represent a brief summary.

The problem started in approximately week 7 of the 18 week course. At that time cheating on tests was observed for the first time (see observation record K). Some weeks later, immediately before, and immediately after the Easter vacation approximately 15 students were still only working through the early units. Circular letters were sent advising them to catch up, and asking them to see the teacher to explain their slow progress. Amongst this group of students, a number were notably lively and vocal
within the electrical engineering department. Feelings began to rise against the course as these students began to realise that they would not be able to complete the course, yet they saw other students (albeit a very small number) who were going to complete the course by cheating. These students also complained about the set book (Elton (1971)) saying that it was too advanced for their use as a single resource, and about the lack of in-course and departmental tutoring help.

In the second week of the second term, three of these students confronted the teacher with these problems in a Monday class period. They made their points about the difficulty of learning from books and the need for more tutorial help quite moderately, but were not satisfied by the response, although it was agreed to meet them again to discuss the course. The following day a larger group of electrical engineering students produced a document reporting the students' views of the course. They objected to the occurrence of cheating, the difficulty of getting tutorial help, the textbook, and the unfairness of the assessment scheme. This document proved to be major catalyst. On the basis of it two meetings with staff were held, the problems were aired, and solutions suggested and decided upon.

4.4 3. Concluding Remarks on the 1975 Course

In June 1975 Particle Mechanics ended with a list of changes planned to overcome the problems encountered, (these will be described in 4.5). The examinations produced results which were not significantly different from the previous years', although the students who were strongly opposed to the course generally got poor results. At this stage I took a positive role in formulating an explanation for the problems and seeking solutions. These explanations contain some important lessons, and although there was no completely agreed consensus, the following interpretations emerged:

(i) Through small and subtle changes, the assessment scheme became a discouraging and punitive element, rather than an encouraging reward system. There is a clear danger in applying notions from behavioural psychology (Skinner (1953)) to the competitive world
of undergraduate teaching by introducing rewards for desired behaviour, in that, those who were unable to achieve this desired behaviour clearly regarded the absence of rewards as a form of punishment. This was the case both when the slower students were denied the opportunity to take the final examination twice, and when they found that they could not meet the increasingly heavy demands upon their rate of unit progress.

(ii) The small amount of cheating which occurred in the course had a very large effect. It took three forms. Firstly, as the testing room was not closely supervised, it was possible for students to use textbooks or notes for answering the tests. Secondly, it was possible for students to get prior knowledge of the test questions, either by removing a test sheet from the classroom which was rare, (the observation record K provides only one example), or more often, it was done by asking a friend about the questions in a given test. Thirdly, the term paper could be copied from a book. All these practices were rare, but they had a significant effect on the weaker students who were also suffering from the first problem.

(iii) The students found learning this course, particularly from the set textbooks, difficult. The difficulty started with the introduction to the course which students found abrupt (L). They felt that a more gradual introduction to the course, and advice on how to work through it was necessary. The textbook presented the next problem. In previous years, the evaluation had shown up shortcomings in it, but in 1975 it was particularly criticised. Added to this was the fact that there was no alternative to the textbook for those students who learnt best by listening, or by using a range of books.

(iv) Students found it difficult to get tutorial help. There was a considerable gap between students and staff in their attitude towards this. The staff felt that they were available for general tutorial help anytime during class, but they would not offer; they had to be asked. The students found it hard to ask for help, felt that the staff were not readily available, and that they were more interested in marking tests. Furthermore, the general tutors in the home departments usually had little experience of the subject of this course.
In reporting the 1973 course it is important not to overplay the difficulties. So finally it should be made clear that the majority of staff and students felt that Particle Mechanics ought to continue to be taught by the Keller plan the following year (1975 - 1976), with suitable changes to return it to its earlier format. These changes were agreed, and the major home department (electrical engineering) agreed to step up its co-operation with the course by providing a staff tutor who would ensure that general tutorial help in Particle Mechanics was available.

4.5 1975 - 1976 The Final Year of the Case Study

4.5 1. Changes Back and Forward

One important task in modifying the 1975 - 1976 course was to return to the more successful format of 1974. The expected minimum number of units was put back to eight, and the idea of holding two examinations was dropped. As a result of a wish to ensure that cheating did not reappear, the single mark credit for passing each of the eight units was dropped, but the deadlines remained, intended to provide a strong guide as to the minimum expected work rate. Also to avoid cheating, it was decided to issue either form of the unit test sheets at random, rather than keeping one version for first-time testing, and one for second.

However, again in 1975 - 1976 changes and developments also took place to carry the course forward. Some had a familiar mixture of aims; partly to solve perceived problems or needs, and partly to continue the development and educational experimentation on the course: Thus the course units were completely revised for the first time. They also included for the first time study advice notes on topics such as learning from textbooks, solving mathematical problems, and note taking. This study advice, and a more general three week series of introductory lectures aimed to help students accustom themselves to the task of individual study. Students had also found difficulty in studying the course because of the single recommended textbook, and for the 1975 - 1976 course it was planned to use a newly established resource centre. Audio-cassette recorded talks
on each of the units were kept there so as to provide an alternative means of learning the course. The resource centre also contained the laboratory experiments. It was located near to the teachers' office, and the offices of other staff able to offer general help on the course.

Undergraduate student testers were used again, although their work and the classroom as a whole was to be more carefully supervised than in the previous year. The branching pattern of units was retained in its simplified form such that freedom of choice was available only to the fastest students although all had some freedom as to the order in which they worked.

4.5 2. 1975 - 1976 Outcomes

It was most important that the 1975 - 1976 course should run smoothly and be seen to have got over the difficulties in 1975. Most measures indicated that this was the case. For example, the results of a general course questionnaire (Q) showed that the students were evenly divided between those who wished to see more Keller plan courses, and those who did not. Student interviews (U) confirmed the course's general acceptability. The presence of a staff tutor from the electrical engineering department had a substantial impact, not only on the acceptability of the course to electrical engineering students, but also on the level of cooperation between the course and the home department.

However, the difficulties of the previous year had continuing effects. The daily record diary kept by the evaluator (R) showed that students had heard bad reports about the Keller plan course from their colleagues in the second year, and clearly the course had to live these down.

One outcome which was carefully monitored was the effect of the changed assessment scheme. Attendance and unit progress are charted in figures 4.1 and 4.2. It can be seen that the unrewarded deadlines continued to have a strong effect on the students, although the pattern is confused in the second term by the fact that the deadlines were put back one week at the last moment, and many students failed to realise this. The overall unit progress was regarded as acceptable, although it is unsurprising that it
was slightly lower on average than when it was rewarded with coursework credit. Finally, the changed assessment scheme and testing procedure resulted in there being no cheating observed in the classroom at all. The questionnaire responses (Q) showed that cheating was once again regarded as an unimportant issue by the students who were quick to point out that it brought no benefits, and that they would resent tight security.

The outcomes of the study advice notes interspersed in the course units were the subject of a special study on the 1975 - 1976 course (T and U). This was conducted jointly with V. Hodgson by means of an extensive questionnaire and interview programme in which 22 of the 61 students were interviewed. Like many other facets of the Particle Mechanics evaluation, this programme was reported in depth to the teacher (Bridge and Hodgson (1976)) and only the major findings are extracted here:

"Of the 22 students interviewed, all had read some or all of the study advice, and most took a neutral to mildly favourable attitude towards it, appreciating its optional nature, and the way it was interspersed within the material. Only one student was strongly enthusiastic about it and claimed it was a great help, while three students had never made use of any part of it. Between these extremes, fourteen students claimed not to have put into practice advice which they had read, but also made statements which strongly suggested they had done so, albeit unconsciously, for example, one student said:

'Usually I have a note book or something, and I try to note down the important points like it's suggested in the book. I suppose that way it helps you remember ...'"

But then he later remarked ' ..... I think it (the study advice) is handy for some people, I don't think I really need it though, but I think it's a good idea to put it there. I mean, you don't have to read it, I don't think it's that useful for me'.

Students with strong positive or negative reactions to the course as a whole showed no particular tendency to hold similar attitudes towards the study advice. There was a marked feeling that the study advice would be most use for learning the later,
harder units.

Amongst the individual items of advice, the audio-cassettes had very small take-up whereas the advice on using objectives for test preparation, had high acceptance, and the practice of note-taking was transferred to and from this course quite widely."

As well as the study advice notes, the audio-cassette unit summaries, and the longer (three week) introduction to the course were also intended to help students who had difficulty in adapting to individual study. The three week introductory period had clear positive effects. Students started the course quickly and enthusiastically, and they had a better grasp of its organisation than previously. The audio-cassettes were less successful. They had been included in the course instead of the occasional relevant live lectures which the students had asked for in 1973, but they did not live up to the promise of the lectures in the introductory period. And furthermore, they had to be used in addition to the book rather than instead of it. The exact cassette usage is difficult to assess as students often forgot to mark down their use on a record sheet in the resource centre. However, the student interviews in 1975 - 1976 (U) suggest that about one third of the class used the tapes, but usually just once, when the students were in the resource centre completing laboratory work.

This points to two main criticisms which students again voiced in the mid-course questionnaire (Q), both of which have been heard since the first course in 1973. Firstly, they found the textbook very difficult as a sole source of reference. Secondly, and possibly as a consequence of this, students wanted occasional relevant live lectures. The audio-cassettes achieved little in supplementing the book, and thus these interlinked problems remained.

Although tutorial help was more readily available for the electrical engineering students in 1975 - 1976, this was not true for the smaller group of physical science students. For them particle mechanics was the only subject studied in a group of more than about fourteen students. And for the first time in 1975 - 1976, they were able to compare this large Keller plan course with a much smaller individual study course taught by the same teacher (Lopez (1976)). As a result they tended to be particularly reserved about
Finally, the use of undergraduate testers became an accepted part of the course, and the procedures for recruiting and instructing them needed little attention. This was also true of the branching scheme. Students continued to value the freedom of choice in particle mechanics (as well as that of pace), whilst at the same time the branching scheme was as straightforward as possible.

By 1975-1976 the examination results certainly could not be used to compare the Keller plan course with the previous lecture course results (Elton (1973)), as a great many changes in the content and format of the course had taken place. Thus, examination results were primarily of use to gauge student learning, and to point out parts of the course which were proving easy and difficult for the students.

The preceding sections show that the Keller plan course in Particle Mechanics went through a number of changes between 1973 and 1976, and many lessons and generalisations emerge from the evaluation. A discussion of these outcomes of evaluation is postponed however, as they also depend upon the aims and the methods of the evaluation itself which are discussed in the next section. Conclusions on both the evaluation and its subject appear at the end of the chapter.

4.6 Evaluating Particle Mechanics

4.6 1. Introduction

Sections 4.2 to 4.5 attempt the somewhat awkward task of bringing together in a single report the many diverse products of four years' evaluation of a complex curriculum development. In doing so many details of the evaluation are compressed and summarised to the extent that it becomes hard to judge the insight or appropriateness of the evaluation itself.

This section tackles this problem by describing the aims, the roles and the methods of evaluation applied to the Particle Mechanics
course. It is not claimed that this section enables assertions to be traced back to raw data. Rather it describes the ways in which the data were gathered, selected, and analysed in order to expose which measures were used to gather the data, when and how often they were used, and what was their return rate, or duration, or format, or focus.

4.6 2. The Evaluation Methods

This section moves backwards from evaluation practice to its rationale, as this represents most accurately the chain of events in this, and probably many other evaluations. For it was only by spending some time investigating such a complex organization as a university course that it was possible to appreciate the aims that evaluation might serve.

This account of the evaluation methods is based mainly upon tables 4.1 and 4.2. Table 4.2 summarises the role and the brief of the evaluation as it changed from year to year, and table 4.1 (p.4/7) summarises the corresponding evaluation methods used, their response rate and their format. They tackle the problem highlighted in section 4.6.1 because each data-gathering instrument is coded (in capital letters), and this code has been used throughout the reports of the evaluation in sections 4.2 - 4.5. Thus it is possible to trace back from the evaluation method to the evaluation finding. It is also possible to move forward, since whenever possible in the table, the evaluative materials (such as progress records or interview schedules) are referenced.

There is one further evaluation technique which does not fit into tables 4.1 and 4.2. This is the evaluation of student learning on the course. This was not a major feature of the evaluation task in any of the four years described, although it was given somewhat greater emphasis in the first year, 1973. Those evaluations of student learning which have taken place were not based upon specially constructed tests conducted either before or after the course, but simply upon the results of a final examination for the course, set and marked by the teacher. In the first year of the innovation, it was reasonable to inspect and compare the examination results on the Keller plan
<table>
<thead>
<tr>
<th>Year</th>
<th>Evaluator</th>
<th>Evaluator's Responsibility</th>
<th>Evaluation Brief and Particular Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973</td>
<td>L. Willoughby</td>
<td>Joint roles of evaluator and course administrator played by research student. Evaluation informal and not intensive.</td>
<td>General Evaluation Brief: &quot;To get a general view of the Keller plan .... and to find out how effective it was in its present context .... and secondly to give feedback on the courses to the course organiser to help improve the courses&quot;. (Willoughby and Boud (1973)). Particularly: to monitor the first stage of an innovation, and assess the joint role of evaluator and administrator.</td>
</tr>
<tr>
<td>1974</td>
<td>W. Bridge</td>
<td>Evaluation was the sole task in the course. No responsibility for day-to-day running of course. Intensive evaluation planned.</td>
<td>General evaluation brief as above. Particular goals: to study the course in depth from an unattached position, and monitor the effects of three major innovatory features introduced for the first time.</td>
</tr>
<tr>
<td>Early 1975</td>
<td>W. Bridge</td>
<td>Joint tasks of evaluator and course administrator. Less evaluation effort than 1974, evaluation was focussed on specific areas.</td>
<td>General brief: to informally monitor the course, and give feedback on the course to the teacher. Particular goals: to focus evaluation on specific issues.</td>
</tr>
</tbody>
</table>
course and the previous lecture course, as well as comparing individual students' Keller plan course results with the results they obtained in other, more traditionally taught subjects. This was done jointly by the evaluator and teacher and is reported elsewhere (Willoughby and Boud (1973)). However, from the second year of the evaluation onwards, the changed assessment scheme, the pattern of optional areas of study, and the introduction of new areas of study meant that comparisons of student learning between the previous traditional course and the developing Keller plan course were of little value. Rather, the examination results have been used formatively, for example, to point out units or areas of the course in which students were under-achieving, to assess the effect of students acting as testers, or to isolate students whose examination results differed markedly from their course-work performance. Indeed this study of examination results was the main feature of the evaluation which was done by the teacher himself.

Course evaluation involves more than selecting, gathering and analysing data. This data must be reported to the teacher in a usable way. In the evaluation of Particle Mechanics, four methods of reporting were used. Most obviously, the evaluator produced lengthy reports on all aspects of the evaluation at the end of each course (e.g. Bridge (1974)). Also to be expected was informal reporting in the day-to-day contact between evaluator and teacher. Two further means were developed. Firstly, from 1974 the evaluator produced a series of interim evaluation reports at approximately three weekly intervals. These performed a variety of tasks including the provision of detailed feedback (for example notes of an error in a test), and the provision of uninterpreted data such as student progress or questionnaire rating results. Interim reports were also used for pointing out emerging issues, and negotiating interpretations of evaluation data between the evaluator and the teacher. In the last year (1975 - 1976) the evaluation was also reported in monthly meetings attended by all the staff of the course. These served a similar purpose to the interim reports, but they also
enabled the evaluation to be quickly disseminated amongst the staff, and gave them the opportunity to either question it, or if necessary take action on the basis of it.

The way, and the extent to which reports of course evaluation are backed up by evaluation data depends primarily upon the reporting procedures and the audience for which the evaluation is intended. In the case of Particle Mechanics, the teacher was the major audience, and the reporting procedures described above were continuous and intimate. Thus interim reports (whether informal, in meetings, or on paper) were rich in primary data, and during the course, evaluator and teacher worked together to develop interpretations of these data. As a result annual reports contained only data summaries, and depended heavily upon the interim dialogue. This approach to course evaluation is fully argued in chapter 5 ("The Process of Evaluation"), where it is included as one element in the theoretical development of an approach to small-scale evaluation called "Microevaluation".

4.6 3. The Role of Evaluation in this Course

The evaluator (myself in all but the first course) was not contracted to perform an evaluation of the Particle Mechanics course with precisely defined goals. At least initially, the compass of aims which could have been pursued was very large. Indeed, the role of the 1974 evaluation owed little to explicit pre-specified aims, or to negotiated values. However, it is clear that the broad context of the evaluation, and the pre-conceived notions of evaluation held by all those responsible for it had a powerful effect on moulding its implicit goals, some of which ran throughout the three-year period. The context of the evaluation was described in 4.1 2. In particular, my position as a research student without subject specialism but committed to a curriculum development project is important to the role I adopted.

What aims did this context prescribe for the evaluation, and what aims were inhibited? Firstly, many factors dictated that the evaluation should focus upon the Keller plan itself,
rather than, for example, the clarity of the subject matter or its pedagogy, relevance, or appropriateness. Hence, of the changes and developments on the basis of evaluation recorded in sections 4.2 - 4.5, the overwhelming majority were to the technique of running a Keller plan course. Constraints which contributed to this were the difference between teacher and evaluator in their grasp of the subject matter, and the orientation of the sponsoring project and educational technology research in general towards innovatory teaching methods.

The close professional contact between evaluator and teacher was responsible for defining an involved and supportive (Boud (1974)) evaluation role. This was reinforced by the evaluator taking on the task of course administrator during the last two years; a degree of involvement in the course which would not have permitted the undertaking of non-participant "illuminative" stance for example (Parlett and Hamilton (1972)).

Finally, the evaluation was primarily formative rather than summative (the "timing" dimension discussed in chapter 2 part II). This characteristic was less rigidly prescribed than the two above. It resulted from the wishes of the teacher for continuous feedback about the successes and failures of this innovative course, and from the fact that the course was in a state of change throughout the period 1973 - 1976, such that a single summative evaluation of it at any one time would soon be outdated. However, there were opposing pressures demanding certain elements of summative evaluation which sprang from the aims of the innovation. In particular, it was important to assess the extent to which the aim was achieved of developing procedures within the Keller plan which allowed it to be used economically for larger classes.

Thus it is possible to extract the aims which the evaluation of Particle Mechanics pursued from the constraints upon it, and the decisions about evaluation made by teacher and evaluator. The evaluation aimed to assist the development of the Keller plan method on this particular course by means of detailed and involved formative evaluation, and also to assess the extent to which changes in the method adopted from year to year indicated
solutions to problems associated with using the Keller Plan in other similar settings in higher education.

4.7 Conclusions

In many respects the previous section on the role of evaluation on the Particle Mechanics course closes this chapter, but there are certain findings of general importance, and links with other chapters which bear repeating.

Of the findings relating solely to this chapter, the lessons drawn from the difficulties with the course in 1975 are perhaps most important. Pages 4/16 and 4/17 discussed the results of changed assessments and patterns of reward on the course. They emphasised that the social setting of a university course plays a powerful role in moulding students' attitudes and their learning, and that there is a danger of applying notions from behavioural psychology to teaching without considering also their social effects. Clearly, assessment too, has a profound effect on students, and subtle changes in the assessment system on an individual study course can have major, and unintended, outcomes. Also, the unfamiliar and unusual nature of an individual study course makes it particularly "accident prone". It attracts attention, being seen as a separate experiment which is not above criticism in the way that the majority of traditional courses are.

More generally, the course has gone through a number of changes since it was started in 1973, with interesting results. The branching pattern of units seems to represent a workable alternative to the more common linear pattern, and it may be appropriate in other individual study courses. The use of students as testers became an accepted part of the course once the need to prepare and supervise them properly was realised. Also, the evidence suggests that the procrastination problem can be overcome if some minimum pace is suggested to students, even if there is no credit to be gained by keeping to this pace.

There are many links with other chapters. Points emerging from the literature survey (chapter 2) and the course survey (chapter 3) concerning, for example, the changing aims of the teaching innovation and its evaluation; the trend of diversification in the methods used;
and difficulties with the meaning of mastery, have all been repeated and extended in this case study. In each case the analysis of these points made in chapter 3 has been largely confirmed.

This chapter also prompts some important questions which the later research tackles. The first relates to course evaluation, for the pattern of research described in 4.6 differs from any single approach discussed in chapter 2/II. In chapter 5, I analyse the kind of focussed evaluation involving close contact between evaluator and evaluated which is illustrated in this case study. On a similar subject, there are questions about the interplay between the innovation and its evaluation, and the effect of the evaluation upon teacher, students, and courses being studied. This is discussed in chapter 6. Lastly, some rather more specific questions about individual study courses remain unanswered by chapter 4. One of them relates to student workload on such courses, which each chapter has so far tackled inconclusively. A second concerns Keller Plan test/tutorial sessions which appear to be both a most successful aspect of the method, and the most controversial when undergraduate tutors are used. Thus finally, at the end of the thesis, 6.2 and 6.3 focus on these two questions, and suggest some preliminary answers.
CHAPTER 5 : THE PROCESSES OF EVALUATION

5.1 Introduction
5.2 The Early Evaluations
5.3 Evaluation by Teachers
5.4 Evaluation for Teachers
5.5 Conclusions
5.1 Introduction

So far the research reported in this thesis has described the two extremes of my involvement with individual study courses: long-standing and detailed research into one course in Chapter 3, and a detached look at all possible courses in Chapter 3. This chapter considers the middle position. It describes three strategies which I have used to provide a tangible amount of formative evaluation for many of the individual study courses in existence at the time of the research.

Of the three strategies, the first and earliest is described only briefly. This was my direct, although limited, evaluation of seven individual study courses (listed in Table 5.1) in the first 18 months of the research, and some of these results have already been drawn upon in Chapter 3 Section 5. However, this first stage provided the basis for a more substantial second stage, involving the research and development of a resource for use by teachers wishing to evaluate their individual study courses themselves. The third strategy took perhaps the opposite stance to the second, for towards the end of the research it was possible for me to engage in direct evaluation on a small scale once more. I noted in Section 2.9 (page 211/19) that most evaluation literature and theorising has concerned relatively large scale projects. By the end of the period of research, I was able to distil from the early evaluations (5.2), and the continuing case study evaluation (Chapter 4), a novel theoretical approach to small scale evaluation, and a description of this is the third part of this chapter. The evaluation resource for teachers was called "Monitorkit" and the approach to small scale evaluation, "Microevaluation"; and these two concepts occupy most of the chapter.

5.2 The Early Evaluations

The early evaluations took place primarily in the first year of research, and were small in scale and formative in purpose. They are listed in Table 5.1. In each case a detailed evaluation report was provided for the teacher's use, and in some cases this was used in published accounts of the courses.
TABLE 5.1

<table>
<thead>
<tr>
<th>Course Title</th>
<th>Teacher(s)</th>
<th>Institution</th>
<th>Source of Published Account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introductory Maths</td>
<td>D. Laurillard</td>
<td>South Bank Polytechnic</td>
<td>-</td>
</tr>
<tr>
<td>Wave Mechanics</td>
<td>P. Allen</td>
<td>Nottingham University</td>
<td>Allen (1977)</td>
</tr>
<tr>
<td>A.C. Theory</td>
<td>P. Black</td>
<td>Birmingham University</td>
<td>Black (1977)</td>
</tr>
<tr>
<td>Quantum Mechanics</td>
<td>L.R.B. Elton</td>
<td>Surrey University</td>
<td>Boul et al (1975)</td>
</tr>
<tr>
<td></td>
<td>B. Stace</td>
<td>Surrey University</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P. Maas</td>
<td>Strathclyde University</td>
<td></td>
</tr>
<tr>
<td>Introductory Chemistry</td>
<td>M. Freemantle</td>
<td>South Bank Polytechnic</td>
<td>Freemantle (1976)</td>
</tr>
<tr>
<td>Vibrations and Waves</td>
<td>S. Kay</td>
<td>Royal Holloway College</td>
<td>Kay (1977)</td>
</tr>
</tbody>
</table>

5.2 1. The Changing Aims of the Early Evaluations

In order to research into an innovation, it must first be established, so the purpose of the first evaluations was support and feedback for teachers unfamiliar with the task of teaching individual study courses. An example of evaluation with this purpose was on the Newtonian Mechanics course which was one of the earliest to be established in the U.K. It was evaluated from outside for the first two years (in the first year this was carried out by L. Willoughby). The evaluation consisted of class observation, interviews with each of the students, questionnaires, and analysis of records of progress. The purpose of the evaluation was clearly formative; changes and modifications were made to the course regularly on the basis of the feedback, whilst summative evaluation and the judgement as to whether to carry on with the course was left largely to the teacher, to be made mainly on the basis of examination results, and his own satisfaction with the course. (This division of evaluation responsibilities is discussed in Chapters 6 and 7). The evaluation of two other courses (those by D. Laurillard and M. Freemantle) also had aims in this area.
As the number and experience of individual study courses grew, different evaluation aims emerged. The first was a research aim, in that evaluation was used to provide evidence for educational experimentation which could be generalised from the single course in question. This first took place on an established individual study course, Quantum Mechanics. This course had previously received evaluation with the first category of aim (support through formative evaluation). The results (only) of the experiment in course transfer between Surrey and Strathclyde Universities were described in 2.4. The aim was to research the effectiveness of the two degrees of transfer (between teachers and institutions), assess the strengths and weaknesses of course transference, and compare the resulting pair of courses. The purpose was not formative evaluation, indeed no detailed changes were made to either course on the basis of evaluation data, and personal support was not really necessary.

The final aim for the research during the initial period was to examine ways of supplementing and extending formative evaluation done by teachers themselves. As has been noted earlier, the first few individual study courses each had the opportunity of outside evaluation. However, courses starting later (in 1974 onwards) did not have this facility. In certain of these courses (those by P. Allen, P. Black and S. Kay) the teachers themselves decided to undertake a modest programme of evaluation, and the aim of my outside evaluation changed markedly. The aim became one of developing, supplementing and ultimately researching this kind of evaluation by teachers, and this resulted in my participation in the evaluation only if and when it was felt that the teachers' evaluation could be extended by outside help.

5.2 2. The Early Methods of Evaluation

Like the aims of the evaluation, the methods also changed and developed. Evaluations with the aim of feedback and support were carried out by means of classroom observation, questionnaire and interview methods, and the study of student progress, test, and examination data. Often study of the latter data was carried out in co-operation with the teacher, and the methods in general involved close contact between teacher and evaluator. During this period, the content and format of questionnaires used in the evaluation of individual study courses was refined and the methods of classroom observation systematised.
Here, as at other points in the thesis, providing evidence is
difficult. One concise and tangible example of the refinement process
to questionnaires exists, however. This is the taxonomy of possible
issues relating to individual study courses, which was generated from
the series of questionnaires I produced. This taxonomy is actually used
in Monitorkit second edition which is described in 5.3, and appears inside
the back cover of the thesis. One test of the completeness of such a
taxonomy is the extent to which it can usefully direct other people's
evaluations, and this taxonomy has been used in many such evaluations,
including one in Australia (Avey & Northcott,(1975)).

In the later stages, the evaluation methods were characterised by a
drop in teacher-evaluator and class-evaluator contact. For example,
the research evaluations on the transferred course were carried out solely
by means of a questionnaire survey, and analysis of the student progress
and examination performance.

Finally, the methods of evaluation aiming to support and supplement
teacher evaluations were different again. Typically, I undertook an
analysis of the teacher-designed questionnaires, and interpretation of
their results. I charted student progress and any other source of
evidence about the course prior to a typically one or two day visit to
the course. Whilst there the evaluation concentrated upon individual
student interviews, because these were found to be difficult for teaching
staff to undertake, and also because they tended to provide key
information which the teachers' evaluation had indicated a need for. For
a teacher's view as to how this process worked in practice, see Allen

5.2 3. Review: The Results of the Early Evaluations

For the teachers involved, the early evaluations resulted in
detailed feedback about their courses which helped them to survive and
develop. It also provided them with moral support, and the opportunity
to learn from the experience of other courses. For me, the evaluations
first enabled a refining of the evaluation methods used, and a chance to
assess their flexibility for different courses and institutions.
Secondly, the later teacher evaluations provided insights into the
processes, strengths, and weaknesses of evaluation by teachers.
5.3 **Evaluations by Teachers**

5.3 1. **Introduction**

The previous section showed that as the number of individual study courses grew, the possibility that each one could receive evaluation by an outsider like myself reduced considerably. This had an important and positive result: teachers were attempting systematic evaluations of their courses themselves. An examination of this was important, for evaluations by teachers have many advantages, most of all they are realistic. In the long run university teachers themselves are the most valuable and most readily available resource for evaluation and research into their teaching. Also the potential relevance of their work and its acceptability to their peers is considerable. Therefore I decided to: research into teacher evaluations of individual study courses; examine ways of developing them; and describe in case study form the actual process of course evaluation by a teacher.

As the first step of the research, a pilot study was planned. This involved collecting together all the possible components of an evaluation approach to be used by teachers, presenting and describing them to a sample of the target group of science teachers, and asking them to select those components which would be most suited to them in thinking about course evaluation. They were further asked how they would wish to see the procedure disseminated and what degree of structure and prescriptiveness they would find most useful. The results (described below) indicated that the procedure should take the form of a loosely structured resource package of printed materials containing plenty of sample materials such as questionnaires and progress sheets.

It was decided to develop such a package in a number of stages. A tentative edition would be produced, disseminated to a manageable sample of the target group who would test it and report back pending a revision of the package after approximately nine months of use. The main thrust of the first edition would be to establish a procedure which would be acceptable and usable to teachers, since considerable danger was foreseen in producing material too obscure and inapplicable to be of use to practising teachers.
The rationale for trials of the second edition differed from that of the first. Once an acceptable and usable package had been produced, it was important to assess the effectiveness of evaluation by teachers using the package. For this reason, I worked closely with two of the teachers using the second edition, examining the evaluative materials they produced and the analyses and data that resulted. Written reports were also obtained from all the teachers about their experiences using the package.

5.3 2. The Monitorkit Programme and the Products

The time scale for this programme was a period of twenty-one months, shown in the diagram in Table 5.2.

The products were of two kinds. First of all there were the two evaluation packages, which are called Monitorkit first and second editions. Monitorkit second edition appears in a pocket on the back cover of the thesis. The whole of Monitorkit first and second edition are to be regarded as documents supporting this thesis.

Secondly, there were reports, descriptions, and conference papers (e.g. Bridge, 1977). These will not be described, but much of the theme which follows derives from them. The essence of the Monitorkit project is a twenty-one month programme in response to an evaluation need, and an evaluation package developed on the basis of research.

5.3 3. The Monitorkit Programme in Depth

This section adds more detail to Table 5.2.

Stage 1: September–December 1974; Background work was performed in a number of areas:

(i) The literature was surveyed for reports of previous research in this area.

(ii) Accessible members of the target group of teachers in higher education and also professional evaluators were contacted and consulted as to what form of assistance they would expect to be most useful.
<table>
<thead>
<tr>
<th>Stage</th>
<th>Period</th>
<th>Material</th>
<th>Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sept-Dec 1974</td>
<td>Collection of possible material</td>
<td>Dialogue with target group on basis of possible material. Discussions of format, content, degree of prescriptiveness and structuring.</td>
</tr>
<tr>
<td>4</td>
<td>June-Sept 1975</td>
<td>1st edition package</td>
<td>Getting overall feedback from teachers who had used package, also written reports, copies of evaluation materials, results, and reports. Gathering commentary from educationists and evaluators.</td>
</tr>
<tr>
<td>6</td>
<td>Nov 1975-April 1976</td>
<td>2nd edition package</td>
<td>Use of package in over 50 courses. Again, gathering detailed feedback from users and educationists. Concentrating on effectiveness of process of evaluation by teachers who were using the package, rather than its acceptability to teachers. Study in depth two particular applications of package, one in ambitious programme, one more modest. Production of descriptive report of progress of teacher evaluations.</td>
</tr>
</tbody>
</table>
(iii) Prior to the programme I had worked alongside teachers performing evaluations. Records of the nature of the help given, and the guidance and materials which proved most productive were studied for indications of a future direction.

The questions posed in each case were these:

(i) What format of advice and help would be most productive and usable?

(At an early stage the option was open to use a range of printed formats, or to convey the advice some other way such as in meetings, workshops, via tape-slide, as a course, and so on.)

(ii) What techniques in the field of evaluation would be most suited to evaluations by teachers?

(iii) What stance should the advice take? Should it present one pre-tested technique for evaluations by teachers, or was some degree of freedom desirable?

(iv) What preconceived notions did the teachers have about evaluation, what procedures did they expect to adopt, and how could this be developed?

No research was to be found on applying the accumulation of experience of evaluation to practising teachers, especially university teachers. Several works on evaluation had attempted to widen their appeal to school teachers (Ten Brink, 1974), but by and large they simply presented the range of study of the professional evaluator in terms simple enough for teachers to understand, with no attempt being made to fit the approach to the needs and problems of evaluation by teachers themselves. Teachers in the target group were clear that they needed a flexible, non-prescriptive form of advice. They saw the need for this in two areas. They wanted to gather subjective reports on the course in question as accurately and systematically as possible. Also, particularly those quite new to the innovative teaching method wanted advice about more objective measures like the significance of examination results, the meaning of student progress, and the monitoring of procrastination. Finally, it seemed likely that a print resource was needed, since much of the advice sought was for sample questionnaires, record sheets and other items of paperwork.
Stage 2: December 1974-January 1975; Production of First Edition of Monitorkit. On the basis of stage 1 a printed resource package called Monitorkit was produced. During the writing stage, teachers in the target group were contacted, and a group, initially of 6 teachers undertook to use the package and provide reactions on it, examples of the procedures, materials and results which it generated. The first edition was not to be widely distributed, it was expected to have numerous shortcomings, and so only teachers prepared to co-operate in the research were included in the programme.

The resulting first edition of Monitorkit was looseleaf, and comprised a one page introduction to the package followed by four separate removable sections. Each section covered various areas of evaluation by teachers which the early study had indicated potential for. Some areas were speculative or supported only by a little evidence, for example, evaluation using student essay/reports, and these were put in to test their potential for general relevance. Each section attempted to distil out the vast body of expertise in, for example, questionnaire design or interviewing, a short descriptive essay which would be relevant to teacher evaluations. Each section explained the approach, the alternatives within it, suggested its strengths and weaknesses when applied to evaluation by teachers, and described the kinds of evaluative data the approach could provide. Where relevant the section also contained sample material (such as questionnaires), a description of how to analyse the resulting data, and short accounts of the previous use by teachers of the approach.

Thus Monitorkit first edition presented a range of evaluation approaches for the teacher to select from and try. It did not prescribe an optimum evaluation scheme or give much advice on how a scheme should be formed, nor did it prescribe the questions that evaluation by teachers should tackle.

Stage 3: January-June 1975; Use of Monitorkit (first edition) in Target Courses. The finished first edition was used by a group of 6 growing to 11 science teachers in higher education. At first it was planned to introduce the package to potential users in a descriptive paper briefly outlining each section, and asking them to select those sections they wanted to receive and use. This proved unacceptable to the users as, without seeing the sections it was impossible to make an informed choice, and the teachers tended just to select sections having immediate appeal (such as those on questionnaires and tests).
From January 1975 to April 1976 Monitorkit itself (in first then second editions) went through a process which it is helpful to think of as evaluation carried out by the present author. In stage 3 this was predominantly formative evaluation, with detailed changes being made to Monitorkit as feedback was received from the teachers using the package. This feedback was of two kinds in stage 3. Firstly, each user had been asked to provide notes on their initial reaction to the package when they had first read it. Eight out of a possible 11 reports were received and many of the changes this feedback suggested were incorporated into the package immediately, by changing the master copy, as the expected total number of copies were deliberately not printed all at once. More overall comment was listed, to be used in the major revision planned for stage 5. The second kind of feedback came when Monitorkit began to be put to use. Then different difficulties arose, and by keeping in close contact with teachers using the package, studying their evaluation plans and materials, and occasionally visiting their courses, I obtained a profile of these.

The picture which emerged was that the evaluations by teachers were strongly based on the package, and often they used the sample materials and the suggested procedures with little alteration. However, two problems were clear; the evaluations tended to be aimless, with no underlying scheme for collecting the data, and also the approaches were selected from the total range of alternatives often with little planning and thought.

Stage 4: June-September 1975; Reporting on Monitorkit (first edition). At the end of the academic year, teachers using Monitorkit (first edition) were contacted again, and were asked for a short report on their use of it, and samples of any evaluative material or reports they had produced on the basis of it. At the same time a group of 4 educationalists and professional evaluators were sent a copy of the package, and asked for notes and comments on it. Rather wider ranging formative evaluation data was collected in this way, and some of the themes which had started to emerge in stage 3 were reinforced.

The reports showed that most of the teachers who received Monitorkit (first edition) had carried out some form of evaluation on their individual study course, usually closely linked to the sample material, and suggested procedures in the package. Thus the package passed the first hurdle, it was read.
Once put into action, problems were reported. The main one was related to the absence of any indication of how to develop an evaluation strategy (or indeed, much mention of the need for one). An illustration of this came in one teacher's report which said:

"This package shows how to ask questions, but not which question to ask, nor how to go about pursuing and answering questions in a logical manner."

Secondly, when I examined the evaluations which teachers had planned, a further problem became clear. This stemmed from the fact that teachers found it difficult to select the most appropriate evaluation method for a given purpose and context. A simple example was one teacher who had failed in his attempt to assess the weaker students' difficulties with the course at a group discussion/evaluation session which students had not been warned to prepare for. As a result the session was dominated by a minority of able and vocal students. Beyond this was the fact that, for example, several teachers did not appreciate the importance of the student progress records in combating procrastination.

Thirdly, having carried out an evaluation, often for the first time, many of the respondents had extremely practical advice to offer, and they felt that this should be included in the revised edition of Monitorkit.

The educationalists' feedback differed from that of the teachers in certain respects. They, too, were concerned about the lack of structuring and the exceedingly non-prescriptive stance. However, they were also concerned that the package presented a limited view of evaluation; that wider reading was necessary to grasp its complexity; that students' views of a course were treated as paramount; and that little advice was presented on judging the success of the course overall, in terms of student learning.

Stage 5: September-November 1975; Revisions for Monitorkit (second edition). The second edition of Monitorkit was based on evaluation findings about the first edition such as those described above. For while the first edition was in some parts speculative trial material, all of the significant changes that went into the second edition can be traced to feedback resulting from the earlier edition. Examples of improvements to Monitorkit through formative evaluation are given below.
The most straightforward improvements resulted from direct recommendations for change. Thus one teacher's finding that evaluation interviews with a class representative could be particularly helpful when he was primed by reading other evaluations of individual study courses was included into the relevant part of the second edition, while the educationist's recommendation that the teachers be given a broader view of the topic was taken up in a descriptive essay introducing the package and a list of further reading at the end.

More important and more difficult to act upon were three clear problems which teachers had found in Monitorkit (first edition)—(i) to develop a strategy in a particular case for asking questions, (ii) to choose appropriate evaluation methods, and (iii) to obtain from Monitorkit generalizable guidance on overall (summative) evaluation of student learning. All three points were tackled in a completely new section in Monitorkit (second edition). The first edition consisted merely of a one page introduction followed by nine separate sections each describing a different evaluation tool. The second edition was in three parts. Part I was a lengthened introduction which went into detail about course evaluation. Part II corresponded to the bulk of the earlier edition containing separate sections on different evaluation tools. Part III "A Strategy for Asking Questions" was quite new, and tackled problems (i), (ii) and (iii) fully, in its major section "A Strategy for Monitoring". This described the various aims of evaluation and suggested how different aims could be matched to different evaluation methods (see the sample Monitorkit 2nd edition inside the back cover). In this case the research basis was both theoretical and practical, for example, drawing upon Scriven's formative/summative distinction and upon experience that showed which methods of evaluation produced formative 'feedback' most effectively. Problem (i) was also tackled directly in a section called "The Questions to Ask". This was developed out of the taxonomy referred to earlier and was intended to help teachers systematise the issues their evaluations would cover.

Other changes in the package are worth mentioning. Part II was changed and developed in the light of experience, length was saved by removing sample material which was not used, and both the well established sections (such as that on questionnaires) and the more tentative sections (such as that on essays/reports on the course by students) were augmented
with examples and advice based on the use of the first edition. As stated, Part III was an attempt to provide help with structure and rational choice, which had been missing from the first edition. It was possible to apply the teachers' experience using the first edition to make the above detailed changes, for example, the straightforward idea of using a single loose-leaf binder to plan and record an evaluation strategy came from one user of the earlier package.

Stage 6: November 1975-April 1976; Use of Monitorkit (second edition). The research programme for the second edition was different from the first. As it was not planned to continue to survey every application of the package and to get detailed feedback about the package in order to make small changes to a regularly used master copy, a larger single production could be undertaken which was distributed more widely than the first edition. A small number of the users were then studied in depth, in order to attempt to get a rather more overall view of the process of course evaluation by teachers.

By April 1976, 32 of the packages had been received by teachers running innovative courses, and the main thrust of research was to study in some detail two selected applications of Monitorkit (second edition) in real situations (see 5.3.4 and 5).

To review, in April 1976 the Monitorkit programme had achieved the following. It had grown out of researches and surveys around an unexplored field, and produced a resource package that responded to the need for evaluation in innovative courses. This package had been put on trial, improved and then used again, and short case-study research was used to look at the actual process of evaluation by teachers. The two case studies are reported in 5.3 and 5.35. In each case I maintained contact with the teacher during his use of Monitorkit, received and examined his evaluation plans, materials and analyses, and discovered what changes (if any) were effected on the basis of the evaluation. This involved class visits and discussions with the teachers as well as documenting their evaluations. *

*Here I would like to thank M. Widden and M. Brown for their help and co-operation.
5.3 4. **First Monitorkit Case Study: A Modest Trial**

5.3 4.1 Introduction: Martin Widden teaches introductory mechanics to engineering science students in the second year at Lancaster University. Since 1973 he has used the Keller Plan (Widden, 1977). He evaluated the course from the start, and for the past two years has used Monitorkit first and second editions in the evaluation. This short study attempts to illustrate the process of evaluation by teachers, and the way it was assisted by Monitorkit.

This course is unusual in this country in that the idea of using students as tutors (or proctors) has been taken up fully, and they have become a vital part of the course as it has grown from thirty students to fifty. Run in the autumn term, the first course (in 1973) was evaluated without Monitorkit, the second was run during the early stages of the first edition, and the third course was evaluated using the second edition of Monitorkit.

5.3 4.2 History of the Evaluation: In 1973, course evaluation primarily took the form of a general course questionnaire at the end of the course. Notes were also kept of aural comments made by students in discussing the course with the teacher. The response rate on the questionnaire was 24/32 and it was mainly rating scales, and response selection questions. Student progress was not systematically monitored, but the examination scripts were looked at carefully to assess changes in the standard of performance.

In 1974 the questionnaire was revised and used again with a response rate of 24/27. It remained similar to the 1973 version, being comprised mainly of response selection questions, but now followed by more spaces for 'further comment'. It was also shorter, clearer and better organised. No records were kept in 1974 of the informal discussions between teacher and students about the course as this seemed to the teacher rather "infra dig", but systematic graphical monitoring of individual and group performance throughout the course was instigated. The questionnaire analysis was similar to that the previous year, revolving around numerical analysis, and again examination results were carefully inspected.

The 1974 course evaluation just described was influenced by the present author who was at the time researching ideas for Monitorkit (first edition), whilst the 1975 evaluation was done using Monitorkit (second
In July 1975 M. Widden sent a written report on his reactions to the first edition of Monitorkit, and an extract appears as a supporting document to the thesis. This report is a good example of the formative data upon which the research and development of Monitorkit is based.

For the autumn 1975 course, the recorded informal discussions with students were reinstated, and organised into a planned scheme as suggested in Monitorkit. The progress record system was carried on as from 1974 also, and the examination results were again carefully scrutinised. The questionnaire underwent major change, and the relevant Monitorkit sections developed both the format and the coverage. The result was that it was far less numerically based than previous questionnaires, more open-ended, and its results were reported to be of more direct use to the teacher. This questionnaire had a return rate of 29/48.

The Effect of the Evaluation. The teacher was surprised to find that the students were glad to be asked for their opinions on his teaching. He also came to value their responses to evaluation, particularly the open-ended questions in the later questionnaires. The evaluation had resulted in a range of developments in the course, and two examples follow.

Firstly the 1975 questionnaire showed a problem over proctor recruitment. It emerged that good students who did not want to be asked to act as proctors sometimes completed relatively advanced unit tests, but held back from having them marked so that they would not be asked to proctor. The solution was to assure students of their right to decline the invitation to proctor, even if they were very well ahead.

The second effect illustrates the danger that teachers may over-react to evaluation data. In the first questionnaire (1973) students indicated that the first four units on the course were revision for them. Consequently the teacher decided to drop the tests up to unit 5 for 1974. This produced several difficulties: the students reached the first test (for unit 5) and failed it because they did not fully understand units 1-4. However, the students felt that they had got through the first units adequately without testing or proctoring, and in the 1974 questionnaire they reacted against the introduction of tests and proctoring in unit 5.
Thus in 1975 tests were reintroduced to the first four units and this aspect ran smoothly, although again, the students stated in the 1975 questionnaire that units 1-4 were revision. Many other changes on the basis of evaluation could be described; some appear in Widden (1977).

5.3 The Effect of Monitorkit. Two points are clearly made. First of all, the effect of Monitorkit was not to introduce evaluation to this course for the first time, since it had taken place prior to the programme. Secondly, its effects on the evaluation programme were modest. It reinforced approaches already being used, it showed a means of systematising an evaluation technique, and it legitimized approaches which the teacher had considered, but regarded as unrespectable. The following extract from the teacher's report makes this point:

"Much the most useful feature, to me at any rate, is that the kit suggests several ways of evaluating a course which I have hitherto either not considered at all or have felt to be for some reason not respectable. Essays, interviews by the teacher, group discussions with the teacher and tape recordings are all in one or other of these categories, and to have the fruits of experience ... is very valuable."

Details of the effects of Monitorkit in this course are to be found in records of conversations with the teacher, copies of the notes and plans for his evaluation measures, and in questionnaires, progress records and other evaluative measures themselves (available as a supporting document). From this material three brief illustrations are drawn:

1. Writing the questionnaires. Over the period of three years, the questionnaires evolved. One major change owes little to Monitorkit; that of the proportion of open-ended questions which were increasingly included due to the relatively useful nature of the data they produced compared with the numerical scales. The content and the format of the questionnaires was modified by Monitorkit though; it broadened the range of questions asked, and made the teacher aware of the wider range of issues which might have been tackled. It also gave the idea of fitting the questions into logical sections, and showed how the framing of the questions could be improved.
2. Discussions with students. Discussions or interviews with students were an example of an approach which the teacher found systematised and legitimised in Monitorkit. He had informally discussed the course with the first group of students in 1973, and kept notes of these. In 1974 this lapsed, but the teacher was encouraged to take up the practice again in 1975 by the relevant section in Monitorkit. By using the general approach suggested, he was able to make notes on the discussions and use them to develop themes to be tested in the 1975 questionnaire.

3. Progress charting. Monitorkit, in the formative stages of the first edition had a direct effect on the way in which the teacher charted the students' unit progress. It provided samples of graphical means for showing individual and group performance, which were used on the course, reducing procrastination.

5.3.5 Comments on Case Study. This case study is representative of many of the applications of Monitorkit. It had modest tangible effects such as improving a questionnaire, or introducing a progress recording technique, and there are indications from the overall evaluation plan that it also had such intangible effects as alerting the teacher to the need for a logical system for evaluation. It also indicates that the awareness resulting from the course evaluation produced worthwhile developments in the course, and avoided possible pitfalls.

5.5 Second Monitorkit Case Study: A More Ambitious Application

5.5.1 Introduction. The 'A' level Physics course at Methodist College, Belfast is an ambitious innovation amongst individual study courses, which contrasts strongly with the Engineering Keller Plan at Lancaster. The innovation will ultimately involve all the staff of the Physics Department headed by Dr. Martin Brown. Computer management of the course is being developed in collaboration with H. McMahon at the New University of Ulster, funded by the National Development Programme for Computer Assisted Learning. Both this funding agency and the staff involved place particular emphasis on evaluation, and for this reason M.B. acquired a copy of Monitorkit, and decided to base the evaluation of his innovation around it.
The innovation is being introduced in several stages. For the academic year 1975-1976 only one of the five 'A' level syllabus blocks (Materials and Solids) is individualised, and use of the computer is restricted to analysis of results and progress. In 1977 all the syllabus will be individualised and computer managed. This case study reports the use of Monitorkit in the first trial of the individualised Materials and Solids block by M. Brown and B. King, a lecturer in Physics at Stranmillis College of Education who was appointed to evaluate this innovative course.

5.3 5.2 The History of the Evaluation. Full details of the course and its management are available elsewhere (McMahon et al, 1977). The evaluation started quite late in the Monitorkit programme, indeed few evaluation plans had been made before the second edition was available. Furthermore, the history of evaluation is rather short since it has been carried out only once, on one fifth of the total Physics syllabus. However, in this period Monitorkit (second edition) has been extensively used on an ambitious course by an outside evaluator, and for this reason this is included as a case study.

The teacher (M.B.) and evaluator (B.K.) worked closely in planning the evaluation of this course, and there was little sense of evaluator and evaluated. Far more than in the first Monitorkit study, in this example, the evaluation was very firmly moulded around the rationale, methods and procedures appearing in Monitorkit (second edition).

5.3 5.3 Applying the Techniques of Monitorkit. The evaluation scheme was planned in four parts. Firstly, a general course questionnaire was devised for the independent learning block on Materials and Solids. This had been done in the two previous years for the Physics lecture and tutorial classes, and the 1976 questionnaire was a development of these earlier questionnaires. This development was affected by reference to the taxonomy of potential issues cited in Monitorkit (second edition).

Secondly, the idea of using feedback slips with the new material was adopted. They were added at the end of the unit tests, and used on three of the core of five units in the block. They were designed using the sample feedback slip in Monitorkit, and included both response selection questions suitable for eventual computer analysis, and open
questions. Thirdly, the potential for gathering student progress data and for record keeping was recognised, although it is probably true that the impending computer management facilities acted alongside Monitorkit to encourage evaluation in this area. Rates of student progress, the paths taken through the optional and remedial branches of units and student attendance were all monitored by a course manager in preparation for the use of the computer. Finally, individual and group discussions with students on the course were also planned for, to be conducted by the evaluator (B.K.) towards the end of the course. Hence, looking at Monitorkit (second edition) it is clear that techniques from Part II sections 1, 2, 5, 6, 7 and 8 were all put into practice on this course. Monitorkit was also used in the analysis stages. For example, the analysis of the free response questions on the general course questionnaire was completed along the lines in Monitorkit (second edition) Part 3.

5.3 5.4 The Aims of the Evaluation and the Effect of Monitorkit. The above section might be taken as implying that Monitorkit solved most of the major problems in the evaluation of the innovatory Physics course at Methodist College. In fact it did not. When I visited the course shortly after it started for discussions with the teacher and the evaluator there emerged a problem over the aims of the overall strategy.

Between them, the teacher and newly appointed evaluator had found it relatively easy to select the data collection methods from part II which would be feasible on their course. They had also successfully completed early stages of the processes outlined in Monitorkit (second edition) part III for starting to develop an overall evaluation scheme. They had decided that the evaluation of the 1976 course should be formative in nature, and have the dual aims of helping the course to survive and improve, and of pointing out approaches which would lead to a worthwhile formative evaluation scheme in the following year when the innovation would grow to cover all five syllabus blocks. However, they found that Monitorkit failed to bridge the gap between these two general aims, and the kind of specific aim or objective which, on its own suggests a question or issue to be tackled, and in a group indicates the suitability of one particular data gathering method rather than another.
illustration of this problem was that they had examined the taxonomy of issues related to individual study courses in part III of Monitorkit (second edition), but were unsure which of them should be pursued in their evaluation.

The problem has a parallel in the transition of broad general teaching aims into detailed objectives by individual teachers on the one hand, and course teams on the other. It is apparent that in the first Monitorkit case study the informal and individual nature of the work allowed this transition in evaluation aims to take place intuitively as it usually does with an individual's teaching aims, whilst at Methodist College the co-operative and open nature of the evaluation produced problems in the same way that course teams have difficulties in agreeing upon specific objectives (Lewis, 1971).

Initially the problem appeared insoluble; the staff were asking for a definition of, or at least priorities amongst detailed evaluation aims for a course which I had little first hand experience of. After lengthy discussions the problem was recast. Generally, the evaluator and teacher were seeking direction on product aims; aims that would directly point out questions and issues. However, it was possible to provide process aims. Hence it was possible to lead teacher and evaluator away from the broad aims of the evaluation towards the issues to be tackled by the data gathering measures (which they had already chosen) by describing a number of processes used by evaluators. By describing the process of focussed development, of producing a portrayal, and of triangulation (see 5.4), the bridge between broad aims and detailed evaluation issues was formed.

There was a further example of the theme which emerges above, of Monitorkit being successful in assisting the largest and the smallest evaluation decisions rather than the intermediate processes. This involves the overall evaluation strategy for the programme which had been successfully generated from Monitorkit, in three parts as follows;
Like the first two stages of the evaluation, the teacher and evaluator had based their tentative plans for the summative evaluation around the notes in Monitorkit. However, once again although the broad aims of the summative evaluation had been decided by B.K. and M.B., the resulting procedures did not clearly follow on. Thus they found difficulty in deciding whether to use a comparative study with students taught traditionally elsewhere, whether to use 'A' level results or results of tests of their own construction, or whether to simply monitor changes in examination performance as the innovation went through its stages of introduction. This was not resolved until I next visited the course when these possible evaluation processes were keyed to the aims they were most likely to satisfy, the degree of effort required, and the possible sources of error were clarified. (See Brown (1977) for the resulting summative evaluation scheme worked out between the Belfast staff and the present author). Again, in the Belfast evaluation, "keying" or "bridging the gap between" general aims for evaluation, and the procedures which might achieve them proved difficult for teachers using Monitorkit.

5.3 5.5 Comments on the Case Study. This case study has omitted some of the detailed features of the Monitorkit-supported evaluation at Methodist College, in order to illustrate the fact that this kind of research programme can throw up evidence of the strengths and weaknesses of the package and the approach. As a result of this research, two improvements to the package appear necessary. Firstly, Part III must place more emphasis on the processes by which broad aims can be linked to detailed evaluation objectives. Secondly, it may be helpful to spell out the range of possible summative evaluation procedures open to a teacher using an innovative method.
Further outcomes of this case study might have been predicted. Monitorkit appears to be used rather differently when it is applied by several people in an evaluation team as compared with individual use. Furthermore, its use by a group tends to show up its deficiencies.

Finally, it is clear that in this instance the present author's direct engagement on several occasions had a marked effect on the evaluation. This may be due partly to the fact that the evaluator appointed (B.K.) started with little expertise in evaluation; but the overall conclusion is that even in the use of a further revised version of Monitorkit, teachers would benefit from occasional contact with a professional educationalist or evaluator, particularly in the planning stages. This need not be a drawback, for experience has shown that this form of co-operation between practising teacher, and educationalist, can be far more productive than the relationship of 'experimenter' and 'experimented' on.

The whole Monitorkit research programme prompts a number of generalisable conclusions about small scale course evaluations by teachers. However, these are postponed until after 5.4 which is a primarily theoretical attempt to expand the role of the evaluator, as opposed to the teacher, to cover the widely ignored area of small-scale educational evaluation (see page 211/19).

5.4 Evaluation for Teachers

Ideally, at this point in the thesis there should be a breathing space. The major theme, and the majority of the work has been described. But as is often the case, there grew out of reflection on the research new ways of looking at the past, and new ideas for the future. This latter part of chapter 5, and the whole of chapter 6, are devoted to the tentative ideas, and the preliminary theoretical and practical developments which have grown out of the main body of the research.

Before launching into this section, it is worth recapping on the theme of evaluation which has been described so far. At the start of the research my evaluations brought me into close and often supporting contact with almost all of the relatively small number of university science teachers who were running individual study courses. As the number of
such courses grew there was both the opportunity and the necessity to change the roles which I adopted. As a result my research became more distanced from the courses themselves, concentrating both on teacher evaluation, and the survey reported in Chapter 3. But in a sense, this progression is reversed in the next sections. For, if Monitorkit is one answer to the problem of what one can do when there is no separate evaluator to help the teacher of an innovative course, then what follows is a re-examination of what can be done if, at least in the first place, there is an evaluator, perhaps put there by a curriculum development project.

The reason for this re-examination appearing towards the end of the thesis is not simply that only after conducting the research programme was it possible to stand back and assess the methods used. A further reason was that at the end of the period of research there was just a little time to turn back to more direct research in the hope of answering a series of key questions which had recurred, unanswered throughout the period of the work. Thus the following argument refers back to the early evaluations such as those in the case study (Chapter 4) and to those mentioned briefly in section 5.2 of this chapter, and it also refers forward to certain detailed evaluations which took place right at the end of the period of research. In describing this I face a problem faced by many researchers in educational evaluation. There is a problem of bridging the gap between any given approach of evaluation, and the day-to-day tasks required of evaluators in performing the evaluation. I have attempted to solve this problem in two ways. Firstly, as the description unfolds I shall use as examples tasks which I performed in the course of the research. Secondly, in the next chapter on the outcome of evaluation, the results and the limitations of an application of the approach are described. However, it must be admitted that this part of the evaluation research has not been taken as far as the work on Monitorkit, and therefore substantial application of the approach has yet to be undertaken, and in this lies a further role of the next chapter, recommending avenues for future research which spring from that reported in this thesis. The following subsections describe Microevaluation; a title which is intended to emphasise its orientation towards the evaluation of small elements of the total teaching and learning scene. It should be noted that the similar term, "micro-evaluation", is used by Rowntree (1974). However, he uses
it to refer to the short term, detailed feedback element of evaluation which I and most other authors refer to as "formative evaluation".

5.4 1. The Need for a Re-emphasis in Small-scale Evaluation

Firstly, it is important to stress that the approach to be developed below uses and extends existing evaluation models rather than create a new model in its own right. It arises out of the problems which I found in applying current techniques to small scale evaluation. However, perhaps the grounding of the problem lies deeper, in the difficult role of the evaluator employed by a curriculum development project, since beyond the idea that I should evaluate the innovations developed by my employing project, little was prescribed.

For the evaluator working with single elements of the total curriculum a further difficulty is that of selecting an evaluation model which has appropriate goals and realistic methods. Often the choice is cast, somewhat unhelpfully, as either psychometric evaluation (covering both Tylerian and comparative evaluation (see Chapter 2 II)), or the more recently developed models characterised by the field research approach of Parlett (1976). Both approaches have been dealt with extensively in Chapter 2II, but briefly; the former, with its basis in disciplines such as psychology emphasises pre-specifying the goals of a teaching innovation, and the devising and administering of tests which assess the extent to which hoped-for levels of, for example, learning or attitude change, have been achieved. The latter has its roots in field research in the social sciences, and places emphasis on educational procedures as well as educational products. In doing so it accepts the goals of the course under investigation as set by the teacher as one source, but not the only source of guidance as to what the evaluation should be concerned with. Rather it responds to issues, goals and problems as they develop and are made manifest during the course.

A technique is needed that will not treat these two models as mutually exclusive, but will both provide data useful for feedback about small scale elements of the teaching programme, and have the flexible and responsive approach characterised by more recent evaluation models. Neither on its own can provide exactly appropriate procedures for the evaluation of small scale innovations, and so an eclectic approach is needed in this situation.
5.4 2. Unfulfilled Needs for Small-scale Evaluation

The three areas in which the selection of suitable evaluation techniques is problematic are described below. I show that the two dominant paradigms demand too much and offer too little, when detailed evaluation of small elements of a teaching innovation is to be undertaken.

(a) Scale. First of all, the need is for evaluation to focus upon small elements of a department's teaching, and a student's learning. However, both existing evaluation approaches are based around larger facets, and in particular that based on field research places much emphasis on the need to investigate the effects of the wide range of forces which exist in the teaching situation upon the subject of the evaluation (see page 211/19). It is assumed that the effects of the individual element under evaluation are so intertwined with those of the course and institution in which it is situated - its traditions, norms, ways of administering and assessing, that the individual element cannot be studied on its own. This notion, and the methods based upon it, have worked well where course evaluation has been conducted (for example Hirschi (1975)), but for the small scale evaluation it means that the detailed focussing is crowded by issues, relationships, and effects quite external to the element under study.

For example, in analysing the interactions during the Keller plan tutorial for the purpose of improving the tests, the most important factors affecting the student are the immediate ones, such as the format of the programme, or his prior knowledge in the subject. In the formative evaluation of these materials, the main interest is in details such as "what kind of diagram makes the point most clearly". An investigation of external factors, such as the student's attitude to the course, would only be relevant if dictated by the analysis of the situation under study. In general, it would not be relevant to the purpose of improving the material. Furthermore, even if data about the entire course were gathered, it might well be that changes and modifications to it would be beyond the brief of the evaluation.
Thus, although small scale events are obviously influenced from outside, the need is first to investigate the subject for clues about the relevant influences, and not to try to follow up all possible influences.

The re-emphasis needed is that it is important not to think in terms of discrete alternative ranges in the scope of evaluation, such as either the tutorials or the course or the entire university scene. Instead, the range should be seen as a continuum; the lower extreme referring to elements such as computer packages or Keller plan tutorials, ranging up through courses and degree programmes, to the departmental climate and eventually to large scale factors such as the personalities of the people and the dynamics of the institution. It is a matter of judgment where the evaluation stops along this broadening range, rather than a matter of principle.

(b) Objectivity. The second problem is with small numbers, taking, for example, a given test or option; the relative insignificance of the one element of teaching in a student's overall consciousness; and the small number of occasions upon which any data can be gathered. An approach is needed that can cope with this real-life situation. However, this is unfulfilled by both psychometric and field research evaluations, both of which depend heavily upon notions of triangulation, cross-checking, and the elimination of bias. However, the low numbers rule out the use of random samples, and since the element under study does not loom large in a given student's consciousness, being only a small fraction of his total course-work, he often has no very strong views on it, and attempts at triangulation must often be ruled out, for repeated questions may tend to force the student to form an opinion he would not otherwise have had. Information cannot always be cross-checked properly.

Given that this is so, it is important to consider the consequences. By accepting a degree of subjectivity and bias it is possible to highlight the complexity of the human situation in which people may not hold a view of the subject at all, or may not be aware of holding one, or they may quite honestly express different views to different people. This fickleness is not necessarily a problem when the aim is improvement as opposed to accurate description. Indeed it is pertinent to question the value of descriptive evaluation where the subject is changing as rapidly as is commonplace in curriculum development.
(c) Orientation. The evaluations required of project evaluators often need to be primarily orientated towards analysis for improvement. Both of the existing evaluation approaches have something to offer here, but more so field research approaches which are often characterised as being "responsive" (Stake, 1974) to the needs of those who are to use the evaluation results, and generally "client centred" (Parlett, 1976). However, there is a need to go further, since evaluators work alongside the teacher who is continuously developing and changing the subject under study. This being so, the evaluation research should at least attempt to yield operationally defined recommendations as well as causal connections, i.e. it must attempt to discover not only why things work or do not work, but also how to change them.

In review, there are three areas, scale, objectivity and orientation, in which there are problems in planning the small scale evaluations demanded of evaluators working with a range of curriculum developments, including individual study courses, which are not met by existing approaches. As a result a re-emphasis which draws on both psychometric and field research evaluation is developed below.

5.4 3. Microevaluation

Microevaluation responds to the unfulfilled needs described above in four main ways.

(a) Probing Outwards. It has already been noted that a difficulty arises in small scale evaluation when investigating a small element of a curriculum. There is a mismatch between the scale of the problem under investigation and the scale of existing methods, so it is important to tailor the methods in such a way that only relevant external factors are followed up.

In a Microevaluation, the initial stages of the evaluation should concentrate on the close observations and analysis of the course element itself, for example the actual interactions in a Keller plan tutorial, transcripts of dialogues, details of interactions with a computer, or written test answers. Back-up measures such as interviews with participatory staff and students, and occasionally questionnaires, would be used to supplement and check the data from the transcripts.
After this initial stage of focussing on the subject itself there is a second stage of developing outwards from this focal point. The strands to be followed up are dictated by the outcomes of the initial analysis. Some such lines of enquiry might appear to merit study at greater depth, while others peter out. To bring this into practical terms, this approach was used in evaluating Keller plan test/tutorials as follows:

The evaluation begins by researching the transactions that actually take place within the Keller plan tutorial session. Written answers and recordings are examined and analysed in different ways, not just by the evaluator but by other tutors, educationalists, physicists, and even students. While some of the transcripts can then be supplemented by interviews with participating staff and students, and some of the issues they raise followed up, this is not necessarily right for all. Suppose, for example, it became clear that one student was responding badly to tutorials because he saw them as a form of assessment. There would be little point in investigating all students' attitudes to assessment to find out why it inhibited them, because that would not help the improvement of tutorials. The time and effort would be far better spent in looking at means for removing this element of stress from the tutorial setting. It is this kind of consideration that will help in making decisions about the relevance of issues to be followed up.

Thus, probing outwards is a realistic attempt to establish all the forces which significantly govern the teaching process in question, by progressively widening the scope of the investigation to concentrate on those, and only those features of the learning milieu which have a significant effect on the small element under study.

The means by which an evaluation focusses upon the subject is important for it prejudges the outcomes of the evaluation to a very large extent. There is some danger that the "probing outwards" of Microevaluation may be confused with the "progressive focussing" of Illuminative Evaluation. In fact they are opposites, as Illuminative Evaluation is very much "macro-
evaluation"; being primarily of value in much larger scale evaluation than the small elements of innovative courses which are the current concern. Therefore, in Illuminative Evaluation the progressive focussing which takes place is generally from the total learning milieu to a small critical element of it. Conversely Microevaluation concentrates initially on the small elements of an innovative course which are the day-to-day concern of the teacher, and only broadens out the investigation to wider issues after this stage. It is important to emphasise again that the two approaches do not compete, they are different strategies based upon experience of different situations, and the experience of progressive focussing in Illuminative Evaluation is now well documented (see for example Laurillard and Dearden (1976)). It is perhaps not surprising that macroevaluations such as Illuminative Evaluation require a strategy for concentrating the investigation which might be thought of as "focussing inwards", whilst microevaluations require a strategy more akin to "diffusion outwards". In practice it is likely that the two stages may follow one another such that, for example, after progressive focussing during a macroevaluation of an entire curriculum, the critical element may be found to depend on further influences in the curriculum, and the study will then probe outwards to account for these. Thus in practical instances, microevaluation may actually follow on from macroevaluation, or vice versa.

(b) Opportunism. Because the goals of Microevaluation are functional, and the number of students is limited, an element of opportunism is essential in small scale evaluation. Instead of random selection, it may be more appropriate to select students who are most articulate, most critical, make most mistakes - and instead of control groups, formative evaluation tends to result in one teaching method being developed as well as possible. In this situation it is important to notice that direct comparison, for example, between an innovative and a traditional method often becomes difficult as the innovation will inevitably develop somewhat different goals from the traditional methods (see 3.5 and 3.6).
The curriculum developer generally relies upon opportunism, and this is particularly true in Microevaluation. This might involve exploiting those situations which produce interesting results, interviewing the most critical students, or using the successes of a few students to indicate how to deal with the others. An analysis of Keller plan tutorials, for example, is not significantly furthered by the transcripts from a silent student. Far more will be learned by carefully selecting those students and staff who will illustrate the problems of tutorials by what they say. The problem of the silent student must be accepted, but tackled in some other way.

Thus the reports from Microevaluation, and the output in general, will be different from the usual research report drawing strongly upon methods developed in field research. The reports will never be final, but work alongside the development of the small scale element, charting its successes and failures. It is most important that the limitations and the methods are made explicit, so that the outsider can judge for himself what the validity of such findings might be.

(c) Continuous development. It is inevitable that this opportunistic approach will occasionally give rise to an unrealistic modification of the material, or an ungeneralisable conclusion in the analysis. Hence it is unlikely that lasting conclusions could be drawn quickly. Further developmental testing and checking would be needed, and this might continue for some time. There is an increasing reluctance in educational technology and curriculum development to see any material as fully developed and static, so such a requirement need not be regarded as anathema. In Microevaluation continuous development is done by the evaluator or the teacher, or both. But it is inevitable that eventually, it would be done by the teacher alone. In planning Microevaluation this has been accepted, and as a result the approach is suitable for the teacher or tutor who does his own evaluation, and the style of evaluation leads towards this.

A key characteristic of this continuous development is that it encourages both teacher and evaluator to make ongoing judgments about the aims which are important for the course at any given time in the study. Again, it is probable that this characteristic is shared with some field research approaches. However, one consequence of the close co-operation of teacher and evaluator, and the orientation towards improvement is not shared with either approach. The result is that course aims can be
revised and realigned in the light of Microevaluation much more quickly than in the commonplace cycle of termly or yearly evaluation reporting. Further, Microevaluation, by concentrating its effort on individual teaching and learning methods, can make use of their short and small scale nature, by providing evaluation feedback repeatedly and often (it might be described as having a short "turn-around" time). This provides the opportunity for repeated judgments of the success or failure of the stated aims, so that development is faster. A formative course evaluation that concentrates on the overall features of the course, and does not break it down into constituent parts for evaluation will not have this advantage, as the delay time for feedback for a course and consequent changes in methods and aims will generally be one year, or at best a term. The Microevaluation approach would be to select one or two features for detailed formative evaluation, and use ongoing judgments of success or failure to determine changes to be made while the course is running. For example, the evaluation of the written coursework units on a Keller plan course could have a "turn-around" time of one week (the time between the first group of students using mark 1, and the second group using mark 2). The only real constraint would then be the time available for rewriting.

(d) Microevaluation is goal directed. The change of emphasis from "finding out how things are" to "finding out what to do" has been mentioned previously. The evaluation methods must indicate what change is needed, rather than simply that there is a need for change. One strategy for achieving this aim would involve examining successful teaching episodes in a Keller plan tutorial, to indicate ways of improving the less successful. Conversely, by concentrating on the critical incidents (Boreham, 1976) in a clearly unsuccessful test/tutorial session, some clear examples of how to improve it may be obtained.

If the goal is improvement, the methods of evaluation will tend to be quite different from those employed in descriptive evaluation for example. The repetition of similar trials will counteract imprecision and ungeneralisability to a certain extent but above all, the methods chosen should be designed to uncover both the problem and the cure.
Microevaluation is a change of emphasis; it is not an attempt to change evaluation models. Rather I have shown that it is possible to derive an eclectic approach drawing upon the two primary models which results in methods and goals that are relevant to the task of small scale evaluation. Further, referring back to 5.1, I have shown that microevaluation and macroevaluation, like field research and psychometrics, should not be seen as mutually exclusive, but should be drawn upon and even alternated between according to the needs of the evaluation. Thus Microevaluation has similar values to field research approaches in that it tries to be responsive, and flexible. It also parallels the developmental testing in psychometric evaluation in that Microevaluation develops as the innovation evolves. It is important to realise that in many cases, not least those cited above, this produces roles for teacher and evaluator that overlap markedly. The change of emphasis is manifested in the four main characteristics of Microevaluation:

1. **Probing outwards** - the evaluation starts by concentrating on the local problem itself, and then develops outwards on the basis of clues and pointers from within the problem.

2. **Opportunism** - data-gathering is opportunistic, and the mode of reporting makes this clear, and benefits from it.

3. **Continuous development** - the process requires continued on-going monitoring, and feedback between evaluator and evaluated in order to counteract undue bias.

4. **Goal orientation** - to a significant extent, the evaluator's focus and methods are dictated by the kind of information which is required by the users of the evaluation, and this involves accepting the need to seek solutions as well as isolate problems.

However, there is perhaps one more fundamental feature to review: the position which Microevaluation takes over the aims of the course or course elements under study. And here it is necessary to return to the analysis of the aims of evaluation discussed in Chapter 2II.

There, it became clear that the two main existing models of evaluation, which are being called psychometric and field research evaluation in this chapter, take considerably different stands on many issues. One important difference is the emphasis which they place upon the stated (teacher's) aims for a particular course or course element, in guiding and circumscribing the task of the evaluator.
Firstly, it was suggested that psychometric evaluation models involve great emphasis upon the stated and predetermined aims of the course, accepting them, and directing the evaluation towards assessing the extent to which the aims are being achieved. Often psychometric evaluations do not attempt to gather data in areas other than those stated in the prespecified aims, nor do they make any judgments about these aims. The methods which follow from this model predictably involve predetermined measurements such as tests primarily, but also attitude scales or observation checklists.

Secondly, field research treats stated course aims differently. Such evaluations attend to the teacher's aims, and examine the extent to which they are achieved; but they are also open to developing further, unstated, but manifestly important aims or unexpected outcomes. Hence field research evaluations treat stated course aims as just one, admittedly important, variable which may mould the resulting methods. Thus, the methods of field research have tended to involve observation, semi- or unstructured interviews, and open-ended questionnaires. However, like the psychometricians, evaluators in the field research would have tended not to take part in judging, revising, or realigning the course aims; this is left for the teacher to do on the basis of the data provided by the evaluations.

Microevaluation contains elements of both. But in the final analysis although it draws from both the prespecified testing methods of psychometrics and the open-minded observation methods of field research, it owes most to the field research model since it is guided both by prespecified aims and by unexpected, manifest, or developmental aims which evolve as the programme is underway. However, there is a further stage since one key consequence of the process of continuous development, and the close relationship between teacher and evaluator in a small and often flexible teaching element is not shared by either major approach. This is the almost inevitable fact that the evaluator becomes involved, with the teacher, in judging and redefining the (usually short-term) course aims, and monitoring the resulting changes in teaching method and learning. This potential for the on-going revision of aims is perhaps one fundamental feature which marks out Microevaluation and may ultimately imply a far closer partnership and overlap of roles between evaluator and evaluated in curriculum development.
5.4 **Implications**

As I have conceived it, Microevaluation is an eclectic approach to dealing with small scale formative evaluation; but some of its characteristic features may nevertheless be applicable to other levels of evaluation. For large-scale evaluations, of a complete course, the course could be broken down into individual components that can be examined repeatedly. On a Keller plan course, the components could be individual units, as these are used repeatedly by different students over a long period. On a laboratory course, the evaluation might look at one experiment out of a "circus", as this could be changed and developed over a period in a way that a whole programme of experimental work could not be. This element of repetition is important for Microevaluation because it is this that allows the re-judgment of aims while the course is running. No matter how large-scale the evaluation, therefore there may always be elements which will be amenable to Microevaluation.

The relationship between the teacher (or developer) and the evaluator has long been a source of debate (see 2.9) and, again, this approach takes a new position. Psychometric evaluation requires only limited contact between evaluator and evaluated, and in field research the evaluator remains distanced to the extent that he presents only his **findings** with little in the way of suggestion for improvement. In Microevaluation, teacher and evaluator would be constantly thrown together to assess and re-assess the short-term aims, and decide how to proceed in future on the basis of this assessment. The two roles tend almost to merge, because the teacher needs to be committed to evaluation if he is to do this, and the evaluator needs some expertise in the subject matter of the course in order to help with such detailed developments in it. The teacher will then no longer be the evaluator's "client", someone who must be treated with care and tact, who must be nudged in the right direction, but never pushed for fear of over-reaction. Instead he would be involved from the beginning in the evaluation strategy, as the evaluator would be involved from the beginning in the design of the technique under study, and the two work together, having slightly different responsibilities, but essentially the same role in curriculum development.
A further implication of this aspect of Microevaluation is that the two roles could, of course, be played by one person, i.e. the teacher could eventually do his own evaluation. With the increasing number of educational innovations, and the decreasing amount of money available for evaluation staff, this is almost inevitable anyway, and the work on Monitorkit reported previously as well as research involving school teachers (Stenhouse, 1975) elsewhere may show the way. Large-scale summative evaluations will always need a specialist, or even a team, but formative evaluation (or "monitoring") in the development of an innovation could often be done by the innovator himself. It should be clear that a logical extension of the use of Microevaluation by an evaluator cooperating with a curriculum developer would be for the developer to eventually take on a modest evaluation role himself, perhaps using Monitorkit.

5.4 6. Conclusions

Overall, the task of section 5.4 has been to explain the ways in which my conception of course evaluation by outsiders crystallised, after a period when I had been for some time engrossed in the problems of helping teachers to evaluate their own courses.

Throughout the section I have been acutely aware of the problems of writing about the practice of evaluation, particularly in isolation from the subject of evaluation. The result is a danger that Microevaluation (in this case, but equally other approaches for other authors) might be seen or imagined as a methodology which is in existence, is widespread, and can be directly taken up and applied. This is, of course, not the case. The title is simply a shorthand for the kind of evaluation I have undertaken and I propose to undertake in the future, and it has been used only by myself and my close associates (e.g. D.M. Laurillard).

My main aim in spelling out the ideas behind Microevaluation in particular, and my evaluation practice generally has been to bring curriculum evaluation down-to-earth, for the strength of Microevaluation as compared with other approaches lies in the fact that it is based upon both research and practical experience.
5.5 Conclusions to Chapter 5

Chapter 5 has dealt with the methods of evaluation in three rather self-contained sections 5.2 to 5.4, each presenting interim conclusions. This section takes these further, and shows how the research on evaluation methods fits into the thesis as a whole. However, perhaps the most important concluding remarks on evaluation appear at the end of Chapter 6 which reports the outcome of using the evaluation approaches described in this chapter.

Turning first to the early evaluations (5.2), it is now clear that they have indeed resulted in a range of established individual study courses, and this has recently been documented in Bridge and Elton (1977). The early evaluations made a significant contribution to my research, and in particular to the planning of Monitorkit. Examples of these contributions have been given. They include the generating of a taxonomy of evaluation issues relating to individual study courses (see Monitorkit 2nd ed, p.50-53), and more widely, providing a picture of the strength and weaknesses of evaluations by teachers upon which the Monitorkit programme could be built.

The Monitorkit programme (5.3) had a second stimulus, as in it I attempted to draw from the great volume of evaluation theorising currently in existence (see Chapter 2II), a core of concepts which would be useful for practising teachers. The test of this is in the package itself, but Monitorkit, particularly the second edition was widely used by teachers.

The idea of teachers performing detailed formative evaluation in a curriculum project in the universities is sufficiently attractive to pose the question: could this approach be used more generally? Reviewing the Monitorkit programme in its early stages, Bond (1976) remarked that success was possible "through the context of the support provided by the wider self-help project, HELP (P)". This does seem near the mark. However, it is not certain that the need for support via a curriculum project is a unique prerequisite. Brown's evaluation of a college physics course (5.3 5) and examples of the use of Monitorkit in, for example, Australia (both outside the influence of HELP (P)) both suggest an alternative: the need for the stimulus and commitment to change which tends to result from an innovatory teaching method.
For the most part the theoretical development of Microevaluation resulted from reflections on Chapter 4. It was developed both as a means of tackling some of the questions which the case study left unanswered (see 4.7) and for analysing the evaluation methods which it used. Essentially Microevaluation is based upon straightforward and rather pragmatic ideas which sprang from my experience of earlier evaluations, and so far the tentative studies to be reported in Chapter 6 present the best evidence about the generalisability of the approach. However, the literature survey described one study more recent than my development of Microevaluation (by Hamilton (1977) (page 2II/14)), which confirms certain important points. Hamilton's notion of "Strong Evaluation" involving "the scrutiny of a given curriculum" and its values parallels Microevaluation's emphasis upon the evaluator's "responsibility for assessing and reassessing (with the teacher) the aims of a course, or part of a course."

In some respects, sections 6.3 and 6.4 contrast, for while Microevaluation is a response to the problem of small scale evaluation when an evaluator is present, Monitorkit responds to the real likelihood that this will only rarely be the case. However, there may be occasions when both approaches are possible, and on the basis of experience with Microevaluation later on, section 6.5 suggests the strengths and weaknesses of evaluation by, and for teachers.
CHAPTER 6 : THE OUTCOMES OF EVALUATION

6.1 Introduction

6.2 The Question of Workload : A Tentative Study

6.3 The Question of Tutoring : A Tentative Study

6.4 The Relevance of Microevaluation

6.5 Evaluation and the Future of Individual Study Courses

6.6 Conclusions
6.1 Introduction

Chapters 3 and 4 have emphasised the first focus of this thesis; the evolution of individual study courses, whilst chapter 5 and the present chapter concentrate upon their evaluation. Chapter 5 described processes of evaluation which were developed and used during the period of research and while many of the findings were reported in the earlier chapters, this chapter goes a little further. Firstly, it describes the preliminary work which took place at the end of the research in extending these findings in two small but critical areas, and examines the contribution of microevaluation in this respect. Secondly, it discusses the fundamental and longer term outcomes of the research. It must be emphasised from the start that the first two sections (6.2 and 6.3) present work which is tentative and preliminary, aiming to suggest paths for future research on the basis of brief initial studies.

It has been said before that after the initial stages of the research, the nature of the course evaluations in which I was engaged tended to become increasingly distanced from the course itself, with the work on Monitorkit being the logical conclusion. However, as well as the continuing contact in Particle Mechanics, this trend was opposed in two further ways; one major and one minor.

The major departure was the theoretical development of microevaluation, described previously. The less major departure was a series of evaluations which I performed at the end of the period of research. These were small in scale, problem centred, and investigative. They were prompted by a number of key questions which the main research had left unanswered. Two such questions emerged. However, in each case little more than preliminary work was done.

The two questions are listed below, and they relate to Sections 6.2 and 6.3, which also discuss why they were chosen in preference to others, and where in the previous research and the literature they spring from.
(A) What is behind the widespread finding that students on individual study courses feel that they require more study time and effort than comparable traditionally taught courses?

(B) It has been established that a major benefit of Keller Plan courses (in particular) from staff and student viewpoints, is the degree of personal contact in the testing situation. What is the nature of this interaction, how does it vary between staff, postgraduate, and undergraduate tutors, and how can it be improved?

All that is claimed of the tentative studies which follow is that they suggest means by which such questions could be more fully answered. There is, however, a further question which emerges from the bulk of the research, about the effect of evaluation upon the future of individual study courses. For it is important to question whether I influenced the innovation I was studying, or whether I had no more effect than Cronbach's (1963) "providing information for decision makers". This last fundamental question is tackled in Section 6.5.
6.2 The Question of Workload: A Tentative Study

Question (A) on the student workload in individual study courses has appeared in a number of places in the thesis. The analysis of significant evaluations of individual study courses in the literature (Section 2.5) showed the importance of the question to previous researchers, and the shortcomings of their indirect or artificially arranged studies. If left the question open, with only untested hypotheses as to why students should rate the courses "hard work", yet not recall excessive time expenditure. The survey of individual study courses, (Chapter 3) also points to a need for research into student workload, as many teachers have made similar findings.

Finally, the case study in Chapter 4 provided a number of stimuli for undertaking this research. For example, in 197^* the questionnaire coded D in Chapter 4 asked two separate questions on workload. The first was "How does the workload on this (Keller Plan) course compare with that on other, traditionally taught courses of equal weighting?" The second asked students to recall how many hours they had spent on studying the Keller Plan course. Whilst the response to the first question was a clear indication that the students felt the workload to be considerably higher than on comparable courses, the data from the second question, when compared with estimates and records of total student private study time, contradicted this.

Many explanations for this discrepancy have been put forward. The most widely held is the belief that students spend a moderate number of study hours on the self-study course, but because self-study is intensive and demanding, the workload and effort seems greater. However, I showed on page 21/37 that although the background literature on this subject is considerable, it is inadequate in one important aspect. All of the literature from both the U.S.A. and the U.K. concerns itself with students' subjective recollection of the length and nature of study periods associated with individual study courses such as the Keller Plan. This data has been gathered using only a limited range of tools such as general course questionnaires and interviews. The only exception was an inconclusive study in the unrealistic setting of a study hall which students were obliged to attend. (Quoted in Kulik et al (1976))
I decided to use an evaluation method not previously used for this purpose, to obtain a direct measure of workload, namely; diaries completed by students at short intervals during a period of days. Thus the aim of this research was to get as direct and objective a view as possible of the study pattern of students on self-study courses.

The study started in the Particle Mechanics course in which twenty study diaries each covering one week were collected from different students over a period of time. The format of the diaries was developed over this period by altering the master copy as experience was gained of the method. One early lesson was the difficulty in obtaining records of weeks which were not grossly distorted, for example by the start or finish of term, interim examinations, and even Students' Rag Week or short holiday breaks. All of these conspired to reduce the number of diaries suitable for analysis to ten.

The results, to be reported below, appeared to merit verification and expansion into a full study. However, the period of research was drawing to an end, and it was possible only to attempt to replicate the results with a further ten students on a different individual study course. Thus in April 1976 ten Third-Year physics students on a Keller Plan course at Royal Holloway College also completed one week self-report study diaries.

They were of a similar format to those used on the Particle Mechanics course, (extracts from the diaries appear in Appendix I/3).

The second pilot study did indeed replicate many of the findings of the first study. However, the small size of both samples (ten diaries in each case) meansthat whilst it is apparent that the process of evaluation developed in this study is valuable, the findings which resulted, although credible and readily rationalizable, are very tentative.
6.2.1. The First Workload Study: Surrey University 1975

The first programme took place in Spring 1975. In all, twenty students were asked to keep a detailed one week diary of their timetabled staff contact and their private study time, on all their courses. Students were given £1 for completed diaries, and this proved to be a welcome and acceptable incentive.

Of the twenty diaries, ten were selected for analysis, the remainder being rejected on grounds of unrepresentativeness or because they had not been fully or seriously completed. The diary shown in Appendix I/3 was designed in consultation with two students working on the course. They advised on the length of the diary and the intervals into which the day should be divided.

The end result consisted of a title page which explained the purpose of the diary and gave instructions for its completion. This was followed by seven identical day sheets divided into hourly periods which the students then further subdivided. These instructions were continuously modified and clarified throughout the period of the data collection but the format of the data record remained unchanged. A notice asking for students to take part in the research was posted, and suitable students were selected from the volunteers.

The diaries selected were numbered 1 - 10, and both the total workload and the pattern and frequency of workload were tabulated. See Tables 6.1 and 6.2.

Throughout this analysis the students' academic workload has been split into three categories; these are "Formal Contact", "General Contact" and "Private Study". Formal Contact time is the total amount of time spent per week in lectures, design and technical drawing classes, and both general tutorials and tutorials on the self-paced course. General Contact represents the formal contact time plus work in the laboratory, and in General Studies tutorials and seminars. (Thus, the general contact is the total time per week in which the student is "taught"). Private study time has been taken as the total time spent by the student on course related study outside timetabled general contact hours. Table 6.1 shows that this study ascertained both the amount of contact time which was set for students, and the extent to which they attended this. Table 6.2 compares the self-paced Particle Mechanics course with the other courses common to the group of students in question.
TABLE 6.1: Formal and General Student Contact Time Ratios, and also Private Study Time Ratio

First Study - Surrey University

(All times quoted in fractions of an hour or minutes)
(P.M. = Particle Mechanics Course Data)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Student ——</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>FORMAL CONTACT</strong>&lt;br&gt;P. M. set</td>
<td>$\frac{3}{17} = 17.6%$</td>
<td>$\frac{3}{20} = 15%$</td>
<td>$\frac{3}{18} = 16.6%$</td>
<td>$\frac{3}{19} = 15.4%$</td>
<td>$\frac{5}{17} = 17.6%$</td>
</tr>
<tr>
<td></td>
<td>Total set</td>
<td>$\frac{3}{26} = 11.5%$</td>
<td>$\frac{3}{26} = 11.5%$</td>
<td>$\frac{3}{26} = 11.5%$</td>
<td>$\frac{2}{24} = 12%$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>P. M. attended</td>
<td>$\frac{2}{16} = 12.5%$</td>
<td>$\frac{5}{17.4} = 4.2%$</td>
<td>$\frac{1}{15} = 6.7%$</td>
<td>$\frac{2}{17} = 11.8%$</td>
<td>$\frac{1}{15} = 6.7%$</td>
</tr>
<tr>
<td>3</td>
<td><strong>GENERAL CONTACT</strong>&lt;br&gt;P. M. set</td>
<td>$\frac{3}{19} = 15.8%$</td>
<td>$\frac{3}{26} = 11.5%$</td>
<td>$\frac{3}{26} = 11.5%$</td>
<td>$\frac{2}{24} = 12%$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total set</td>
<td>$\frac{3}{23} = 13%$</td>
<td>$\frac{2}{23} = 8.7%$</td>
<td>$\frac{2}{23} = 8.7%$</td>
<td>$\frac{2}{23} = 8.7%$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>P. M. attended</td>
<td>$\frac{2}{18.8} = 10.7%$</td>
<td>$\frac{3}{23.2} = 3.2%$</td>
<td>$\frac{1}{22} = 4.5%$</td>
<td>$\frac{1}{22} = 4.5%$</td>
<td></td>
</tr>
</tbody>
</table>
| 5       | **PRIVATE STUDY**<br>Total | $35\frac{1}{2}$ | $12\frac{1}{2}$ | $22.20\text{min}$ | $16\frac{3}{4}$ | $26$
| 6       | P. M. | $9\frac{1}{2}$ | $7.55\text{min}$ | $7.10\text{min}$ | $1\frac{5}{7}$ | $2$
| 7       | P. M. / Total | $26.8\%$ | $63\%$ | $32\%$ | $10.5\%$ | $7.7\%$
| 8       | P. M. Private Study | $9\frac{1}{2} - 1$ | $7.55\text{min} - 2.15\text{min}$ | $7.10\text{min} - 2$ | $1.45\text{min} - 1$ | $2 - 2$
| 9       | Minus P. M. Contract<br>Missed | $8\frac{1}{2}$ | $5.40\text{min}$ | $6.10\text{min}$ | $0.45\text{min}$ | $0$
| 10      | P. M. (General Contact<br>attended plus Private Study) | $11\frac{1}{2}$ | $8\frac{3}{4}$ | $8$ | $3\frac{3}{4}$ | $3$
|         | Total (General Contact<br>attended plus Private Study) | $54\frac{1}{4}$ | $36$ | $44$ | $39\frac{3}{4}$ | $48$

† Including 10hrs general study work  * Including 2hrs general study work
Table 6.2 Number of occasions/week a given subject turned to in private study. Also, total private study times in given subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Particle Mechanics</th>
<th>Mathematics</th>
<th>Computing</th>
<th>Field Theory Machines &amp; Power Circuit Theory</th>
<th>Electrical Measurements &amp; Physical Electronics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>T</td>
<td>0</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>Student No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>9.30</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>7.55</td>
<td>1</td>
<td>.20</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>7.10</td>
<td>5</td>
<td>6.10</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1.45</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5.25</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2.15</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2.40</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>9.15</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.30</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total 0 and T</td>
<td>39</td>
<td>48.25*</td>
<td>15</td>
<td>20.30</td>
<td>15</td>
</tr>
<tr>
<td>Total T</td>
<td>Total 0</td>
<td>1.24</td>
<td>1.37</td>
<td>1.47</td>
<td>2.04</td>
</tr>
</tbody>
</table>

0 = No of occasions  
T = Total Time (hr/min)  

* Total T less time released by non attendance at class periods = 34.10
From Table 6.1 it can be seen that the Keller Plan course (Particle Mechanics) was assigned ("set") on average approximately 17% of the total Formal Contact time or 13% of the total General Contact time (Table 6.1 rows 1 and 3). Comparison with rows 2 and 4 shows that student attendance at the (usually optional) Keller Plan course periods was on average lower than that on the remainder of the courses which were usually compulsory.

The individual study course occupied a wide range of proportions of private study time from 1.7% to 63% of the total (row 7). The most common range (6 out of 10) was 10-30% with two individuals spending proportionally more than this range, and two proportionally less. Thus, comparing row 7 with rows 1 and 3 there is little indication that overall, students spent much more or much less private study time on the individual study course than its weighting in terms of formal set contact time would predict.

This indication can be arrived at by a different means of analysis. Row 9 shows the ratio of the total time spent by students on particle mechanics both in and out of class, to the total time that students spent on their academic study. Using this means of analysis the range of proportions is smaller, but again the average of approximately 14% is again neither higher nor lower than would be predicted from the weighting of the Keller Plan course.

Row 8 was obtained by subtracting from the total Particle Mechanics private study time the number of hours of optional Particle Mechanics contact time which the student did not attend. This was done since students would be expected to do Particle Mechanics private study when they were absent from the optional classes.

Table 6.2 relates solely to private study time. It presents the number of separate occasions a given subject was turned to by students for private study, and the resulting total private study time in each subject over the one week periods. The Keller Plan Particle Mechanics course is compared with: (a) the Mathematics course which has identical lecture contact weighting as the Particle Mechanics course (3 hours); (b) computing which has 1 hour per week contact, and for which private study may be idiosyncratic due to the need to use the computer unit at given times; (c) the three overlapping courses Field Theory, Circuit Theory, and Machines and Power, which total 3 hours per week. The three
combine to form effectively one course with common tutorials and so on; and finally (d) Electrical Measurements, and Physical Electronics again each 1 hour per week courses with overlapping tutorial arrangements.

Thus Table 6.2 compares the private study time of all the main theoretical subjects taken by all the students surveyed. A number of interesting, though tentative, observations can be drawn from it.

(1) The individual study Particle Mechanics course occupied the students for more private study time than any other common theory courses, except the combination of the Field Theory, Circuit Theory and Machines and Power courses. Even when the total Particle Mechanics private study time is reduced by the time released through non-attendance at the optional class periods (See Table 6.1 row 8) the picture remains the same. Students spent less time on courses they would be expected to regard as more important, such as Mathematics and Electrical Measurements and Physical Electronics. At first sight this might seem to contradict the findings from Table 6.1 (that Particle Mechanics does not occupy an excessive proportion of the total private study load). Closer inspection shows that the explanation lies in the fact that courses other than the five common theoretical courses tabulated in Table 6.2, such as laboratory work and general studies occupy disproportionately more private study time than their set contact time would predict.

(2) The total number of occasions on which students turned to Particle Mechanics private study exceeded that for any other course, substantially, and without exception. This is even true for the combination of three courses (Field Theory, Circuit Theory and Power and Machines).

(3) The average work stretch for Particle Mechanics private study was shorter than any of the other common theory courses, as measured by Total T divided by Total O. (This is despite the high Total T noted in (1)).

(4) Particle Mechanics was the only course for which all of the students did at least some private study during the weeks recorded.
6.2 2. Interim Discussion

This study reinforces some established results, but also shows quite a new angle on the problem of individual study course workload. On the one hand, it does show that on average students spent a longer time doing Particle Mechanics private study than private study for any other course, (except the combined Field Theory, Circuit Theory, and Machines and Power Courses). This is true even when the optional nature of the classes in Particle Mechanics are accounted for.

However, the study also draws attention to an aspect of self-study course workload not previously considered. Students undertake private study on the Particle Mechanics course on more separate occasions than any other course, and on twice as many occasions as any course except the combined courses in Field Theory, Circuit Theory, and Machines and Power. Furthermore, having turned to Particle Mechanics private study, students spent on average a shorter interval studying it than any other subject. And finally, Particle Mechanics was the only course surveyed for which every student did at least some private study in the week concerned.

The following explanations for these results are put forward. Self-study course material is always available for study, and as none of the students surveyed had finished the course, the students could make obvious progress by working on it. Thus students finding themselves with free time could always do Particle Mechanics private study, and could do so on a number of occasions at times during each week. However, since there were no real deadlines on this course, students would not be forced to spend the long periods of private study time which tended to occur just prior to the very definite deadlines which existed on other courses. In other words, the continuous availability of self-study material for private study, combined with the lack of firm deadlines for completing it, encouraged the "frequent but short" pattern of Particle Mechanics private study noted above.

These tentative conclusions based upon a small sample have been checked by looking back at the original source, the diaries, and there are indeed a number of reinforcing examples such as students fitting a short period of Particle Mechanics private study into a pattern of other activities.
6.2 3. The Second Workload Study: Royal Holloway College 1976

This second programme took place in Spring 1976. The diaries
used were of the same format as previously used at Surrey University,
although the instructions were changed to suit the different circum­
cstances. The procedure used for the second survey was very similar

to the first and will not be repeated here. Again, the diaries were

numbered 1-10, the data from them was tabulated, and the categories
"Formal Contact", "General Contact", and "Private Study" (defined
previously) were used. Table 6.3 is the result of the same analysis

as 6.1 and similarly 6.4 and 6.2.

Comparison of Table 6.3 and 6.1 shows that the variety of workload
and workpattern between individual students was even greater in the
second study than the first. Re-examining the diaries shows that the
first study involved first-year students at a point in their degree
programme where their work was more thoroughly prescribed than in the

case in the third-year when many more options are available, and when
long projects are often undertaken. This point should be born in mind
throughout.

The wide range of responses means that the average of 40% for the

proportion of set Formal Contact time devoted to the individual study
courses, and the corresponding figure of 32% for the General Contact time
proportion are of limited value. One point is clear, however; this
second study concerned a course given higher priority, at least in terms
of timetabled and (see row 7) private study time than the first.
However, rows 2 and 4 do not repeat the finding in the first study, that
attendance at the individual study courses was on average lower than the
remainder of the courses.

The individual study course occupied a vast range (from 0% to 100%)
of proportions of private study time. And while the average proportion,
35% should be treated with corresponding caution, it does repeat the
finding of the first study that "there is little indication that overall,
students spent much more or much less private study time on the individual
study course than its weighting in terms of formal set contact time would
predict." This is checked in row 9 which has an average of 29%; con­
sistent with the original findings.
### Table 6.3 Formal and General Contact Ratios, and Also Private Study Time Ratio

**Second Study - Royal Holloway College.**

(1 S = Individual Study course in Computer Electronics Design and Data Transmission Physics)

<table>
<thead>
<tr>
<th>Row No.</th>
<th>Student ———</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>FORMAL CONTACT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 S set</td>
<td>$\frac{1}{6} = 16.7%$</td>
<td>$\frac{4}{9} = 44.4%$</td>
<td>$\frac{2}{9} = 22.2%$</td>
<td>$\frac{4}{6} = 66.7%$</td>
<td>$\frac{4}{6} = 66%$</td>
</tr>
<tr>
<td></td>
<td>Total set</td>
<td>$\frac{1}{6} = 16.7%$</td>
<td>$\frac{3\frac{1}{2}}{8} = 40.6%$</td>
<td>$\frac{2}{9} = 22.2%$</td>
<td>$\frac{4}{6} = 66.7%$</td>
<td>$\frac{2\frac{1}{2}}{4\frac{1}{2}} = 51%$</td>
</tr>
<tr>
<td>2</td>
<td>1 S attended</td>
<td>$\frac{4\frac{1}{2}}{9\frac{1}{2}} = 47.3%$</td>
<td>$\frac{4}{9} = 44%$</td>
<td>$\frac{2}{10\frac{1}{2}} = 19%$</td>
<td>$\frac{4}{20} = 20%$</td>
<td>$\frac{4}{15} = 26%$</td>
</tr>
<tr>
<td></td>
<td>Total attended</td>
<td>$\frac{4\frac{1}{2}}{9\frac{1}{2}} = 47.3%$</td>
<td>$\frac{3\frac{1}{2}}{8} = 40.6%$</td>
<td>$\frac{2}{10\frac{1}{2}} = 19%$</td>
<td>$\frac{4}{20} = 20%$</td>
<td>$\frac{2\frac{1}{2}}{11\frac{1}{2}} = 21%$</td>
</tr>
<tr>
<td>3</td>
<td><strong>GENERAL CONTACT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 S set</td>
<td>$\frac{4\frac{1}{2}}{9\frac{1}{2}} = 47.3%$</td>
<td>$\frac{4}{9} = 44%$</td>
<td>$\frac{2}{10\frac{1}{2}} = 19%$</td>
<td>$\frac{4}{20} = 20%$</td>
<td>$\frac{4}{15} = 26%$</td>
</tr>
<tr>
<td></td>
<td>Total set</td>
<td>$\frac{4\frac{1}{2}}{9\frac{1}{2}} = 47.3%$</td>
<td>$\frac{3\frac{1}{2}}{8} = 40.6%$</td>
<td>$\frac{2}{10\frac{1}{2}} = 19%$</td>
<td>$\frac{4}{20} = 20%$</td>
<td>$\frac{2\frac{1}{2}}{11\frac{1}{2}} = 21%$</td>
</tr>
<tr>
<td>5</td>
<td><strong>PRIVATE STUDY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total Formal</td>
<td>62</td>
<td>14</td>
<td>18$\frac{3}{4}$</td>
<td>8</td>
<td>6$\frac{1}{2}$</td>
</tr>
<tr>
<td></td>
<td>Total General</td>
<td>62</td>
<td>14</td>
<td>24$\frac{3}{4}$</td>
<td>8</td>
<td>6$\frac{1}{2}$</td>
</tr>
<tr>
<td>6</td>
<td>Total 1 S.</td>
<td>41$\frac{1}{4}$</td>
<td>7$\frac{3}{4}$</td>
<td>3$\frac{1}{2}$</td>
<td>8</td>
<td>3$\frac{1}{4}$</td>
</tr>
<tr>
<td>7</td>
<td>1 S / Total General</td>
<td>66.5%</td>
<td>56.4%</td>
<td>14.4</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>8</td>
<td>1 S Private Study</td>
<td>41$\frac{1}{4}$</td>
<td>$7\frac{3}{4}$ - $3\frac{3}{4}$</td>
<td>3$\frac{1}{2}$ - 0</td>
<td>8 - 0</td>
<td>$3\frac{1}{4}$ - $1\frac{1}{2}$</td>
</tr>
<tr>
<td></td>
<td>Missed</td>
<td>= 41$\frac{1}{4}$</td>
<td>= 7</td>
<td>= $3\frac{1}{2}$</td>
<td>= 8</td>
<td>= $1\frac{3}{4}$</td>
</tr>
<tr>
<td>9</td>
<td>1 S (General Contact attended plus private study)</td>
<td>$45\frac{3}{4}$</td>
<td>$11\frac{7}{4}$</td>
<td>$5\frac{1}{2}$</td>
<td>$\frac{12}{28}$</td>
<td>$\frac{4\frac{3}{4}}{18}$</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>STUDENT NO.</td>
<td>INDIVIDUAL STUDY COURSE</td>
<td>COMPUTER SCIENCE PHYSICS</td>
<td>PHYSICS</td>
<td>MATHEMATICS</td>
<td>STATISTICS</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>4.1</td>
<td>2.3</td>
<td>10</td>
<td>20.5</td>
<td>2.1</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>7.2</td>
<td>0.9</td>
<td>5</td>
<td>3.5</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>3.5</td>
<td>1.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>8.1</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3.5</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>5</td>
<td>1.7</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>12</td>
<td>2.4</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>5</td>
<td>0.7</td>
<td>11</td>
<td>16.5</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>8</td>
<td>1.6</td>
<td>2</td>
<td>2.3</td>
<td>1.4</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2.5</td>
<td>0.5</td>
<td>2</td>
<td>7.5</td>
<td>3.9</td>
</tr>
<tr>
<td>TOTAL D AND T</td>
<td>57</td>
<td>79.5</td>
<td>67.5</td>
<td>4</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>TOTAL T</td>
<td>140</td>
<td>178</td>
<td>175</td>
<td>164</td>
<td>2.125</td>
<td>1.625</td>
</tr>
</tbody>
</table>
Table 6.4, like 6.2, relates solely to private study time. It presents the number of occasions a given subject was turned to for private study, and the resulting total private study time in each subject over the one week periods. Again, all the "theory courses" taken by the students are represented but the range is far greater in this third-year course, with only one course (Computer Science) nearly as widespread as the individual study course.

The first of the findings from the previous study (that the individual study course took more private study time than any other common theory course) cannot be checked due to this range of options. However, the fact that six diaries showed options which had occupied more private study time than the individual study course tends not to support the original assertion.

The second finding from Table 6.4 on the number of occasions upon which private study was undertaken, presents similar problems. This time the individual pattern gives no guidance, with equal numbers of students who have studied the individual study course on the most occasions, compared with those who had studied some other subject on more occasions. However, the total "0" for all students on the individual study course is certainly much higher than that for any other course represented.

The third and fourth findings can be checked. The average work stretch for the individual study course was considerably shorter than that for the only course of comparable popularity (computer science), and this was also true for all the options with only one individual exception (Operational Research). Finally, with only one exception, all the students did at least some private study on the individual study course during the week of the diary, while there were three students who spent no private study time on computing and one further student failed to spend any time studying the operational research course he was following.
6.2 4. Discussion of the Findings of the Workload Studies

These small scale studies have pointed out a number of new perspectives in the question of student workload on individual study courses. Although the most straightforward question, (that of the balance between the workload in total hours spent on these innovative courses and other, traditionally taught, courses), was not unequivocally answered, the studies do: (1) suggest that the difference is not substantial, and certainly less substantial than indicated by responses in questionnaires, so that the rationalization in terms of the more intensive use of time is probably sound, (2) point out that there are significant variables other than total time, so that the question itself cannot be framed quite so straightforwardly.

A difficulty in interpretation also arises from the fact that priorities which courses should receive are not made clear to students and it is highly probable that they are perceived differently by different students (see for example Miller & Parlett (1972)). A further observation, and indeed difficulty, was brought to light in this analysis, for compared with the amount of class time, practical courses and also general study courses occupy a disproportionately large amount of students' private study time.

Perhaps the newer and more interesting group of findings related to the number of occasions during private study upon which the individual study coursework was turned to, the length of the resulting period of private study, and the apparently regular weekly attention to the coursework. The second study appears to support the explanation for these phenomena put forward in 6.2 2, and this will not be repeated here.

There is more to be said about the outcomes of this small scale evaluation and its relevance to microevaluation, but it is convenient to postpone this until after a description of the much briefer initial steps which I have taken in answering question B.
The research I have done on this question is very much a preliminary investigation for the future, and although fragments of the answer are available, its main value to the thesis is that it represents exactly the kind of question which microevaluation was developed to tackle.

N.B. Probably the most graphic description of this method appears in the videotape (Boud Bridge & Laurillard (1975)) also available in print (Boud Bridge & Laurillard (1976)).

This question focuses on the Keller Plan test/tutorial session itself. The first problem is quite simply of gaining access to the transactions which take place in these sessions. This requires preserving, by means of recordings of some kind, the essence of each test/tutorial. A number of methods were considered, including observation of the sessions by a third party, (the evaluator); asking the participants separately to recall the sessions; using audio or videotape to record the interaction; and examining the test papers and notes referred to during the sessions. The first two approaches were tried, but it was found that the presence of a third party overhearing and noting the tutorial influenced both student and tutor markedly. Similarly, the recollections of the tutor and particularly the student did not convey the richness of the situation, nor the critical points at which, for example, learning was facilitated or blocked. A videotape of the interaction would have been ideal, but its production would have upset the class sessions dramatically. So, by process of elimination, it was decided to attempt to audio-tape record a selection of test/tutorials, and gather the written material discussed in each test.

The tape recording proved to be quite unobtrusive and tapes were obtained from both junior and senior staff members, from postgraduate tutors, and from students on the course who acted as "testers". The research was confined to one course, Particle Mechanics at Surrey University where the general formative evaluation (Chapter 4) had shown up dissatisfaction on the part of some of the course members who were tutored by their colleagues.
A total of 32 recordings of individual test/tutorial sessions were obtained from eight different tutors over a period of just three weeks. It was important to obtain all the transcripts at the same point in the course since it has often been noted that the nature of the interaction changes as the course proceeds, since students become accustomed to defending their ideas and answers in the tutorial situation. This admittedly interesting variable was not one which the evaluation was directly concerned with.

As students wrote their answers to the test questions on the test sheet, it usually proved adequate to keep this as a record of the written transactions during the test/tutorial. The number of units which the student had taken and the point in the course which student testers had reached were also recorded.

6.3 2. The Analysis

Only the initial stages of analysis were undertaken, and there was no obvious single means of doing this; the tape recordings provided at least two possible means of analysis; either directly from the tape, or from transcripts. Also, the written material, i.e. the tests and the students' answers, presented a problem: could it be keyed to the dialogue at any moment in time, and if so, would it add to the analysis?

The preliminary analyses involved a group of 8 University staff members, all of whom had the required knowledge of physics, but some of whom were primarily educationists (listed below). The analyses used the written transcripts of the tape recordings for a number of reasons. One obvious and necessary reason was to preserve the anonymity of the tutors and students. Another reason was my concern not to divulge to the group of outside specialists the status of the tutor conducting each test/tutorial until after they had a chance to react to the dialogue.

Eleven test/tutorials were selected since they covered the range of tutors (from professor to undergraduate), and a range of students (from well above average in terms of unit progress, to well below average). The test/tutorial sessions lasted from 3 to 18 minutes, and thus, if the panel were to be given a manageable quantity of material, some form of sampling was required.
This was done by transcribing approximately the first and last 30-60 seconds of each session, and the middle 120 seconds or so. However, this was not rigidly adhered to. For example, 'beginning' and 'ending' extracts were selected in such a way that they were self-contained as far as possible, whilst starting and finishing at a natural pause in the dialogue. Similarly, on three occasions the middle 120 seconds proved to be of dubious value; for example in one case it comprised of the student talking under his breath as he wrote out a modified answer.

Once the transcripts were produced the plans and procedures which were devised for use with the panel of outside experts were tried and discussed with two university staff of similar background to those in the panel itself.* These trials suggested that the written records, and in particular the test solutions, added little to the transcripts, as it was impossible to relate the two direct. Therefore, I decided not to issue any of the written material to the panel. Subsequent feedback, especially from the physicists on the panel suggests that this was too extreme. The best compromise would have been to issue the exact text of the test questions without the students' written solutions.

An example of one of the set of transcripts appears in Appendix 1/4. The remainder of the transcripts are supporting documents to this thesis. From the example it can be seen that the status of the tutor was not printed on the transcript. Rather, this was listed separately and issued in an envelope.

In order to balance the difficulty which the two staff who piloted the procedure had encountered, one further more lengthy transcript of a "traditional" tutorial was included to provide a baseline and rough comparison.

The participating academic staff were also given a sheet of instructions as to how to perform the analysis of the extracted transcripts. The substance of these instructions was a direct attempt to answer question (B) from a variety of points of view.

* I would like to thank B. Stace and P. Ring for their help in this
I accepted that there could not be one agreed viewpoint as to the nature of any given tutorial, nor of the strengths and weaknesses of a range of comparable test/tutorials. The procedure was an attempt to reflect this diversity of viewpoints, and to derive from the physicists and educationists on the panel their own ways of judging, categorising, and classifying 1:1 teaching interactions. The instructions were as follows:

"(i) Read through the "traditional tutorial" and the test extracts once or twice to get the feel of them...write notes and overall comments on the tutorial transcript as a whole, and the Keller Plan test transcripts as a whole.

(ii)...Specifically, read through the tutorial transcript and the test extracts and write notes which compare the two types. You might look at differences and similarities for example, or strengths and weaknesses...

(iii) Now concentrate on the 11 Keller Plan test extracts. Go through them from 1 to 11 and write down the characteristics of each. As you go through write down what category of tutor you think each is lead by, and say what it is about the extract which makes you think this... Open the envelope (listing the tutors' identities) and note the identity on each extract. Read through the extracts again, and write out your reactions... In particular, say what you now think are the characteristics of the four groups of tutors.

(iv) Summarise your views about Keller Plan test sessions as such, and in comparison with more traditional tutorials, also your views on students as tutors... Then return all your notes."

(Bridge (1975))

The result was a series of eight analyses of the extracts from the panel; M. Cox, D. Hurd, B. Stace, P. Ring, D. Laurillard, S. Kay, B. Hodgson, and B. Barrett, (those by B. Stace and P. Ring being different (as explained), they piloted the method which was changed slightly for the rest of the group in the light of experience).
As stated earlier, at this stage of the research, time was running out, and I only carried out a preliminary analysis. However, this does convey some idea as to the kinds of outcomes an extension of this work would have in answering question B, so a selection of tentative findings are presented below.

6.3 Some Tentative Findings

(The following section is based upon structured responses from each member of the panel. One example of these appears as a supporting document).

In response to instruction (i), half the panel responded that they were surprised by the "imprecision" of the spoken word, or its "crudity" or "lack of depth". Also, those members of the panel who were primarily physicists found it difficult at first to concentrate on the tutorial or the Keller Plan test when they could not be sure exactly what was the test question which was being discussed or written down.

The comparison of Keller Plan test transcripts with that of the traditional tutorial (ii) showed a wide range of reactions. By and large the comparative comments related to features of Keller Plan tests which do not exist in the more familiar tutorial situation. Five of the panel commented that the Keller Plan test sessions were more purposeful and clearly directed to the students' work, with little of the "side chat" that was noticed in the traditional tutorial. The panel evaluated this in different ways, some feeling that the tests were too regimented and strait-jacketed by the test format, and others feeling that the clear purpose of the session outweighed this. Secondly, the panel noticed the rather clear element of assessment, which affected the Keller Plan tutorials. In fact, each panel member observed that in one or more of the extracts this resulted in students failing to divulge difficulties in the test situation, or in them appearing ill at ease (particularly towards the end where the pass/fail decision is made). Finally, in comparison with the traditional tutorial, the Keller Plan test sessions tended not to contain what one panel member described as "Mini-Lectures". This reaction was shared by all but two of the panel but again they evaluated it differently. In particular three people who were primarily physicists found this a failing in the respect that the tutor would not be drawn into lengthy explanations on physics, particularly in areas outside the bounds of the test.
However, the remainder took a different view, feeling that there was more value in "getting the student talking".

A unique comparison made by one of the panel was the lack of continuity in the Keller Plan test situation. She noted that there were never back references such as... "like we discussed last time" or "which you had difficulty with before".

The comparison within the Keller Plan extracts forced the panel to decide upon criteria for choosing between experienced and inexperienced tutors (for they were not told this at first). The following were features of such a session that were widely thought to be desirable:

- The tutor went beyond the question, picking up problems and mistakes rather than accepting them. (This was often termed "probing").
- The session did not appear threatening to the student.
- The tutor was knowledgeable and clear.
- The tutor handled the session skilfully. (one panel member said "firmly in control of the session").

However, the criteria upon which different panel members compared the extracts differed greatly in emphasis. Six of the panel attempted to use these criteria to judge the status of the tutor in each extract. Thus sixty-six such decisions were taken. In total, 60% of these decisions proved correct. Of the remainder, 17% were decidedly incorrect. For example, an undergraduate tutor being taken for a senior staff tutor or vice versa. The panel found it difficult to recognise undergraduates amongst the tutors, and often confused them with postgraduates. Conversely, the senior staff members were quite recognisable, providing the only extract upon which the panel were all correctly agreed. However, it was noticeable that correct decisions on extracts of undergraduate tutoring were made on the basis of two unique criteria; the tutor being sympathetic with the students' problems (in 7 cases), and the tutor making a reference such as "that is what they want" (8 cases).
When the panel compared their judgements with the actual identities of tutors in the extracts, two were strongly impressed with the undergraduate tutors, and two felt that they emerged as unsatisfactory tutors, although they had actually failed to locate them from the transcripts. It was agreed that authority in handling the tutorial tended to increase with status, and that only staff tutors would provide what one panel member called "broad tutoring". Tutors of lower status were seen as having correspondingly lower knowledge of physics. At the level of undergraduate the two panel members cited above thought this was inadequate, whilst the remainder had a reaction such as "adequate within the compass of the test". Once they had learnt the actual identities, many of the panel also noticed the undergraduate tutors "looking for what 'they' want", and sympathising with the student having difficulty getting the "right answer". The postgraduates were often said to be on the borderline between students and staff. Yet apart from characteristics related to status difference (perhaps simply age), their tutoring was described in exactly the same way as the undergraduates.

Lastly, the panel were asked to provide overall reactions to the results of the analysis, and the procedure. The Keller Plan test/tutorial was seen by all as a valuable learning setting, indeed the four panel members with no experience of Keller Plan courses in action were surprised by the fact that far more than marking was involved in these sessions. Two specific reactions come from over half the panel. Firstly, they felt that variations within each category of tutor were as significant as the variations between categories of tutors. Secondly, they felt that some simple training would help the undergraduate tutors, and that they should only be allowed to tutor units they had already built upon by studying later units.

Commenting on the methodology, it was clear that the panel members who were primarily physicists would have found the test questions and solutions valuable. To quote one: "the audio channel is weak, but the written channel may be stronger."
Many of the panel expected wide variations to occur in the way in which other colleagues would evaluate tutorials, and three felt that only by this kind of eclectic study could the range of priorities be accounted for. One particularly interesting idea was that the criteria for "what makes a good tutorial" could form the basis for tutor-training sessions; (this corresponds very closely to the "goal orientation" in Microevaluation, See 6.4).

6.3 3. Training Keller Plan Tutors

The easily produced tape recordings of the Keller Plan test/tutorial could be used live by tutors wishing to develop their own skills in the 1:1 situation; for the transcripts study showed that these are rather different from the skills used in other, more familiar tutoring situations. The procedures that such a "train yourself" approach might adopt are discussed in the videotape (Boud Bridge and Laurillard (1973)).

This idea is important because tutor training on Keller Plan courses emerged as desirable both from the comments received from the panel in this study, and the difficulties experienced during the introduction of undergraduate tutors on the Particle Mechanics course in 1975 (See Chapter 4). Also, it is probable that the panel members in the transcripts study had deep seated reasons for the reactions I have reported, and, for example, some may have felt threatened by the idea of undergraduates acting as tutors. Where this is the case, the facility for listening to recordings of tutoring privately may allay the teacher's concern and help him to develop his own Keller Plan tutoring. Further, tutors could learn from the students' remarks on the tapes the features which make test/tutorial valuable, stressful, frustrating, and so on.

This takes the discussion back to the starting point; gaining access to the dialogue which tutor and student take part in, and focusing on this small but critical element of the Keller Plan.
6.4 The Relevance of Microevaluation

This section looks at how far the tentative studies on student workload and Keller Plan tutorials reflect the ideas about small scale evaluation developed in the previous chapter. However, it is not realistic to try to rigorously relate each feature of Microevaluation to each feature of these studies for two reasons. First, this was not their primary purpose, which was to answer the questions A and B. More fundamentally, in reviewing the literature on evaluation I showed that approaches such as Illuminative Evaluation, or Stake's matrices of evaluation, have mainly been influential in suggesting new emphases or priorities for evaluation which evaluators have interpreted in many different kinds of procedure, rather than in prescribing in detail one procedure to be followed.

The priority in Microevaluation is focussing upon small elements of a teaching method, and I suggested that four evaluation characteristics might follow from this. Of these, three were present in 6.2 and 6.3: probing outwards; opportunism; and goal orientation.

Both the workload and the tutorial studies involved probing outwards from the focal point of the research. The former started by concentrating solely upon students' total private study time for one particular course (taught by Keller Plan), but the study broadened out to consider work periods and even the surprisingly heavy workload produced by laboratory and general studies courses. Similarly, the initial focus upon the transactions in test sessions broadened out to a consideration of the influence of assessment upon them and ways in which tutors could be trained in the art of running such sessions.

There were elements of opportunism in both studies. For example, the workload study relied on the small number of diaries that I could obtain from reasonably co-operative students at a particular point in the courses. More so the tutorial study, where in selecting the extracts for transcription I took the opportunity of using students with revealing attitudes or difficulties.

The test/tutorial study illustrated goal direction, towards understanding and improving this feature of the Keller Plan. The analysis went beyond a description of the events in the session to discover what features made them successful, and how this could be used to help tutors.
Only continuous development was absent from the studies, and mainly because of time pressure it was impossible to use the style of short cycle evaluation described in the previous chapter. However, it was stated in chapter 5 that Microevaluation was not just a plan for the future but also a way of analysing my evaluations in the past. This being so, it is clear that continuous development played an important part in, for example, my evaluation of the Particle Mechanics course (chapter 4). There, my close contact with the course and the teacher resulted in very frequent evaluation reports, revision of aims, and action based upon them.

I suggested above that the important test of an approach like Microevaluation is whether it affects the emphasis of evaluation rather than its detailed procedures. This is clearly so in 6.2 and 6.3. Each section concerns an aspect of individual study courses both vital and controversial in the literature from the U.K. and the U.S.A. Yet I have located no other studies which have concentrated upon the teaching element itself, (tutorial contact or student workload), even in the huge U.S. literature on the subject. Whereas there have been many studies replicating survey findings that 1:1 contact is regarded as "an important feature of the Keller Plan", or that heavy workload is "one of the most widely reported difficulties" (See 2.5). The emphasis of Microevaluation on the value of focusing upon small and critical parts of the total curriculum, at least as the first step, is reflected in both studies. The emphasis has several beneficial effects. For although the results of both must be tentative until a full scale study can be undertaken, they do refine considerably the questions to be asked, and the expectations of the answers. This may be a strength of Microevaluation which could be exploited since emphasis upon the individual student who is being tested, reading, planning his study time, or revising for an examination, prior to the macro-scale studies of the whole learning milieu suggested by Parlett and Hamilton (1972), for example, may provide a useful check that the really important issues are being addressed.
The job of chapter 6 is to look to the future. Sections 6.2 and 6.3 did this by presenting tentative studies which I undertook at the end of the period of research that point to areas into which future work might profitably grow. This section takes a different stance, for it discusses the longer term effects of the evaluation of individual study courses upon teachers, students, and curricula more generally, by reflecting upon the research that I have reported.

The first point to be made is that the nature of the evaluations I carried out during the period of my research changed substantially, and that different kinds of evaluation should be expected to have different kinds of outcome. The changes themselves, although discussed previously, deserve further comment. Firstly, they were an attempt to respond to the needs of the particular curriculum development. It has been argued that the initial need was for formative evaluation of a relatively modest and rapidly changing innovation, and in this area psychometric evaluation has little to offer. So "supportive evaluation" (Boud (1974)) was followed by survey research and then course monitoring by teachers.

Indeed, the research has taken account of the potential contribution of teachers as fully as possible. The package "Monitorkit" described in chapter 5 is the major outcome in this area, and since the end of the programme it has been reprinted, made more widely available, and used by a growing number of teachers. This in itself is an interesting achievement, and much could be learnt about the process of course evaluation by teachers from a follow-up study of the teachers who have received and used Monitorkit.

The supportive and process-orientated stance of the early evaluations and of Monitorkit may have longer-term effects upon the teachers running individual study course. The most notable example of this is to be found in teachers' motivation for taking up individual study courses. Chapter 3 showed that while they started by aiming for cognitive improvements such as those shown by examination results, later on most of them were motivated to continue using the method because they saw it achieving process aims such as getting students to learn independently through reading. And the nature of my evaluations may well have influenced this change, for the methods used place
far greater emphasis upon describing processes than measuring products. In a situation where the evaluator and the teacher work closely on devising the aims of evaluation, it is open to question as to where this priority sprang from initially, but there can be no doubt that it will tend to be perpetuated by the emphasis placed in Monitorkit.

However, the work reported in this thesis leads into areas which would normally be the province of the researcher rather than the teacher. There, the main need is to continue the work just started (and reported earlier in this chapter) into more overall, and perhaps eventually summative evaluations of individual study courses. Extension of the focussed evaluations is just one priority, and summative research through both psychometric and large scale "illuminative" evaluation may also be important.

The results of the workload study have important implications for future overall research on individual study courses, and research into tertiary level teaching in general. It obtained records of students' work pattern in the real situation, and established that not surprisingly perhaps, a teaching method markedly different from that used in the majority of university courses can prompt students to adopt markedly different learning patterns. This highlights the idea that future research should pay at least equal attention in relating the outcomes of, for example, Keller Plan courses, to the learning methods they prompt, as compared with the attention they pay to details of the teaching method.

The results of the tutorial study also have implications for future summative research. The use of a panel of teachers to evaluate the data, and the variety of the interpretations they arrived at, calls into question the value of producing a single portrayal of a course no matter how thoroughly researched. Chapter 211 showed the disagreement between evaluators upon how they should interpret their findings in presenting summative evaluations. Bringing these two points together suggests a more realistic way of helping teachers to decide upon whether or not to take up individual study courses. This would involve presenting teachers with a range of interpretations by other teachers as to the successes of such courses, for them to select from, and form their own summative evaluation. This is the strategy
The general lesson may be that university teachers will accept direct help from an outside evaluator in formative evaluation once they have decided to take up an innovation. But their overall summative decision about adopting the method in the first place is based upon their own interpretation of the reports they see of the method, and their judgement as to how far it fits into their context, their ways of teaching students, their beliefs about the purpose of university science teaching, and so on. And in this latter area, the best the evaluator can do is to provide them with examples of how other teachers have made interpretations such as these. Returning now to the original question as to the long term effect of evaluation on the development of the Keller Plan, a number of points become clearer. The effect is a general awareness of the teaching and learning process on the part of teacher and learner; this much is obvious. However, the range of potential focus is wide, and it is likely that the orientation taken by teacher and student is profoundly yet subtly influenced by the evaluator. The questions which an evaluator poses, the areas of priority which he chooses to research and the range of feedback which he reports all serve to manipulate the way in which teacher and student think about the course. This is clear now only in retrospect, but the effect is plain to see. My predisposition was to study individual study courses as a teaching method, and the questionnaires, interviews, and evaluation reports bear the stamp of this priority. As a result students and teachers I worked with tended to take on the same focus. Developments or faults in the teaching method were attended to, to the exclusion of, for example, updating the curricula or monitoring student learning. This may well have a bearing on the fact that individual study courses at the moment are generally innovative in teaching method, but often traditional in course content. Throughout the thesis I have argued that this focus was appropriate, and that the primary need was for a workable alternative to the lecture. However, it is likely that this emphasis on teaching method innovation occurred for other reasons also. The courses which have been evaluated have been predominantly at an introductory level, and often service courses. Such courses do not have the flexibility
of curriculum or assessment which exists in more advanced courses. Therefore it is possible, but by no means certain, that in such cases course evaluation and consequently teacher emphasis can only be placed upon teaching method as the only realistic variable in this setting.

Heightened awareness of the process of teaching and learning does not have exclusively beneficial results however; there is a danger of over-reaction on the part of teacher or student. Both have been recorded in this thesis. The case of teacher over-reaction was shown in the first Monitorkit case study. There I described how the format of the first few units of a Keller Plan course was changed in response to feedback from the first group of students who found them too easy, only for the succeeding group of students to find the revised scheme too difficult resulting in the format reverting to the original. This was a direct result of the results of a questionnaire evaluation.

The case of student over-reaction occurred in the case study course, Particle Mechanics at Surrey University, and was described, although not in these terms, in chapter 4. The intensive evaluation of the course in the first two years resulted in much more feeling being generated about the course than in the third year when the reduced emphasis on evaluation (along with certain other worthwhile changes to the course itself) resulted in the students seeing the course as a more routine and unexceptional part of their degree programme.

These examples illustrate the point made in the development of Microevaluation that repeated questionning can force students to adopt attitudes and positions that they would not normally take. And the important parallel is that repeated information gathering may force the teacher to take decisions or make changes that he would not naturally make. In individual study courses being evaluated it would be very difficult for students to end up having no opinion of the teaching method, as they are asked about it so frequently.

There is a further outcome of evaluation which concerns the aims of individual study courses. This results from my emphasis on formative rather than summative evaluation and the way this has interacted with the teachers' aims. In this case the effect has been to make teachers' aims open to change, even during the course itself. Thus few attempts to implement individual study courses have been made in a manner which might be described as the thorough trial of a pre-designed programme.
Rather, the teaching method has often changed substantially even during the first experience of the course. Again, this dynamic form of curriculum development seems particularly appropriate during the early stages of a curriculum development.

To conclude this section on the outcome of evaluation, there are two more general observations to be made. Firstly, I have often found it hard to assess the influence of evaluation upon courses, teachers, and students, because of the setting of much of the work reported in this thesis. In particular, many of the teachers who have been involved with my evaluations have also been deeply involved in new initiatives in teaching generally, and their courses have often been part of a complex scene of change and resistance to change. Again, an example is the physics course at Methodist College, Belfast (the second Monitorkit case study), where the course and its evaluation were just a small part of a radical change in the way science was to be taught in a whole college and locality.

The final effect of evaluation may be the most important. For my work has shown that it provides teachers with a legitimate and respectable heading under which they can discuss with one another down-to-earth matters such as how their teaching is going, and also for thinking deeply, occasionally for the first time, about what they are teaching and why.

6.6 Conclusions

There is little to add to the preceding section which draws out most of the important conclusions about the outcomes of evaluations, but some points merit extension or summary.

The chapter started with two tentative examples of small scale evaluation on questions of student workload and test/tutorial contact, and both studies should serve to sharpen future discussion on these issues. Influenced by the emphasis developed in Microevaluation (5.4) their focus was precise, and their methods were unusual amongst curriculum
evaluations. In this respect they go some way to showing how evaluation theorizing can help to frame evaluation plans; a point which is not always clear in the literature.

The first study showed that students' pattern of work on an individual study course is considerably different from that on, say, a traditional lecture course, and went on to suggest other characteristics of undergraduate study. The explanation for this difference (page 6/8) seems important, but there is a further point to draw out. Knowledge about the workpattern which results both from the range of innovative teaching methods now available, and from traditional teaching methods, may provide most valuable guidance for matching together academic subjects, the teaching method used, and the student being taught.

The tutorial study described one way of giving a panel of tutors the chance of analysing themselves Keller Plan tests run by tutors from Professor to undergraduate level. The results showed just how differently this data can be interpreted, and they also suggested extensions of the method into tutor training. This important point about the variety of interpretations was also made in Microevaluation; it was enlarged in Section 6.4, and it may play an important part in planning and reporting similar evaluations in the future.
CHAPTER 7: SUMMARY AND CONCLUSIONS

7.1 Introduction
7.2 The Results of the Research Programme
7.3 General Conclusions
7.4 Further Outlook
7.1 Introduction

This chapter concludes the thesis by describing the achievements of the research reported, discussing their general relevance to the fields of educational innovation and evaluation, and looking forward to future work and the questions that arise. At the beginning of the thesis, page 1/1 described the two subjects as research into individual study courses and the development of appropriate evaluation approaches. The discussion that follows shares this dual focus, but also considers the interaction between the two.

7.2 The Results of the Research Programme

7.2.1 Evaluation of Individual Study Courses

The first task of the research was to analyse a developing range of alternatives to traditional university courses which place great emphasis upon individual study. An important starting point is the predominantly American literature on the subject which highlights widespread approaches like the Keller plan and the audio-tutorial system. Chapter 2 described the growth of methods such as these, and showed the importance to their originators of the bases of these methods in learning theory. The literature survey also outlined the marked diversification which has taken place since the methods were first devised. Analysis of the most significant evaluations of individual study courses resulted in a clear picture of their popularity with students and teachers, despite equivocal reports about their demands upon student time and upon teaching resources. An important result of the analysis was to call into question on two grounds certain widely reported studies of student learning as reported in teacher evaluations (e.g. Kulik et al (1976)). The literature also showed the need for research on approaches to curriculum evaluation both due to the failure of existing methods to answer important questions about individual study courses, and the dissatisfaction of teachers with the outcomes of these evaluations.
The results from a survey of 43 individual study courses in 37 colleges, polytechnics and universities in the U.K. extended the picture produced by chapter 2. I divided these into courses where the material was presented at the students' own pace, and those which were teacher paced, although the latter could be further subdivided into large and often ambitious resource-based courses, and smaller courses which often used group teaching a good deal. The different methods had different results, for example, student paced courses like the Keller plan produced close staff-student contact, accurate records of student progress, and problems with procrastination by the weaker students. The smaller teacher paced courses were fitted into the conventional departmental setting more easily, but group contact replaced personal contact, and students could fall behind with less chance of being noticed.

The survey had certain more fundamental results. In particular it showed how teachers' aims for taking up individual study courses changed, from product aims concerned with examination results and retention, to process aims like promoting independence and book-studying skills. At the same time evaluation changed from measuring products to monitoring processes, and thus it is likely that teachers' aims and evaluation methods interacted, although it is difficult to tell which led and which followed. Individual study courses had made teachers acutely aware of problems of teaching and learning, and the survey described many initiatives and innovations being undertaken by teachers. One such initiative confirmed the trend located in the literature survey; that of diversification. Thus it was suggested that the main value of well defined methods such as the Keller plan was, and still is that they provide a well tried starting point for teaching innovation, rather than an end in themselves.

Three quarters of the courses surveyed had been evaluated, by the teacher alone, or with outside assistance. Usually this involved the study of examination results and a questionnaire written by the teacher. Although the survey found teachers willing to perform evaluation, their results were not as helpful as they might have been, due to their limited experience of evaluating.
One Keller plan course, Particle Mechanics at Surrey University, was the subject of a three year case study, and this added life to the survey results. For example, it demonstrated the diversification of teaching method by studying the effects of a branching programme of course units and options. Even the limited range of choice was regarded as more important than self-pacing by the students, possibly because student paced learning (but not teaching) is in fact a feature of every course! Undergraduate tutors were used on the course; they benefitted from the experience, but the study indicated their need for simple training in order that their tutees should also benefit.

Important results of the case study concerned the effects upon students of manipulating both incentives for student progress, and assessments. At first the two were linked, and it was found that progress deadlines awarded with course-work credit reduced procrastination dramatically. Later, course-work credit was dropped but the deadlines still acted as a strong incentive. Lessons such as these were disseminated to many other individual study courses during my early research, when I briefly evaluated nine other courses. Thus I was able to test the general relevance of findings from the case study.

Results concerning other manipulations of the teaching method were equally important, although some were unsuccessful at the time. They were produced by further attempts to reward rapid unit progress, by giving fast students the opportunity to take an earlier examination. Although these changes were in accordance with Keller's plans, (which in turn rely upon Skinnerian psychology), the result was dissent by students who could not achieve these rewards. The competitive social setting of a university science course meant that these students regarded the absence of such rewards as an unfair form of punishment. Perhaps an equally important conclusion from these events was the accident prone nature of innovations, and the importance of close monitoring.

The case study further illustrated the changing aims of individual study courses, and the influence of evaluation. Evaluation again focussed the attention of both teacher and student
upon the course, but this was not always beneficial, and examples were found of over-reaction by both teachers and students in the major case study in chapter 3 and in a smaller study in chapter 5.

Lastly, two tentative studies concentrated upon questions of student workload and Keller plan test/tutorials; these being issues left open by the main research. The former suggested that the pattern of student study on Keller plan and similar courses is markedly different from that on more traditional courses; explained the differences and possible reasons for them; and went on to propose that this may explain previous contradictory findings on workload. The tutorial study suggested the key characteristics of these important sessions in a Keller plan course, and showed similarities and differences between tutors from Professor to undergraduate, though this study showed how widely different teachers evaluate these sessions, and it indicated how staff and students could be trained in Keller plan tutoring.

7.2 Approaches to Educational Evaluation

Again for this second focus of the research, the literature survey contained an important analysis, in this case showing how the vast range of options open to the evaluator may be systematised by means of six dimensions for decision about evaluation. The literature revealed little recent experience of the type of small scale formative evaluation which I subsequently undertook. For although early approaches to evaluation which formed part of Tylerian curriculum development, for example, did have this focus, their sole emphasis upon measuring intended outcomes to the exclusion of processes meant that they were unsuitable for the purposes of my evaluations. The newer approaches such as those by Stake (1973) and Parlett and Hamilton (1972) do emphasise processes, but in practice their focus is upon large scale summative evaluation. Thus the literature survey showed that research on approaches to course evaluation was necessary, in part to bring evaluation back in a different form to its status in the Tylerian system, where it is sufficiently small in scale to be useful and feasible for teachers.

The survey of courses found that evaluation made teacher and
student pay increased attention to the course, as described above. It also suggested (as did the main case study) that my evaluations of teaching method influenced the process of innovation such that changes in method predominated over, say, changes in content, (although this may also be a result of the level of the courses studied). These evaluations produced on-going feedback which highlighted difficulties as they occurred. I found that one result was that the teachers changed their plans and procedures (as well as their aims) rapidly as a result of this, so that individual study courses have rarely been completely stable. The history of change on the Particle Mechanics course is a good example of this.

Chapter 5 produced most of the results of the evaluation research. It acted upon the needs expressed in the previous two paragraphs for research into evaluation approaches capable of use by teachers, and more generally for evaluation approaches appropriate to small scale formative evaluation. I examined the needs and the potentials of evaluations by teachers. The results suggested that the teachers' commitment to use the results of evaluation was important, and was enhanced by their taking an active role in devising the strategy and methods. Examples and guidance were needed however, and my response was an eighteen month research programme examining the effects of two packages which aimed to facilitate teachers' evaluations. These were called Monitorkit first and second editions, and the results suggested that teachers involved in a curriculum project, or engaged in teaching innovation can develop their teaching by this means. Two short case studies of course evaluation by teachers were described. An important finding was that teacher evaluation must involve an overall strategy, from which appropriate aims, procedures, and data gathering instruments can be decided upon. Teachers found difficulty in this step, and although they can be guided by resource materials, this is one worthwhile focus for outside support.

The Monitorkit programme resulted in a number of data gathering procedures of proven value to teachers, and a taxonomy of issues concerning individual study courses which their
evaluations might select from and pursue. Although the short case studies again illustrated the danger that evaluation may result in over-reaction, they did show how teachers can systematise their evaluation to include procedures they would not normally consider. Finally, by contrasting the case studies, I suggested that evaluation of a course by an individual teacher produces different difficulties, and perhaps fewer difficulties than evaluation of one course by a group of teachers.

Analysis of the evaluation procedures in the major case study of the Particle Mechanics course, and plans for the focussed evaluations on workload and test/tutorial sessions, resulted in a new theoretical approach to small scale formative evaluation which I called Microevaluation. The analysis showed difficulties in existing approaches concerning scale, objectivity, and orientation; and suggested that characteristics of opportunism, continuous development, and goal orientation were more appropriate to Microevaluation. Two resulting emphases are important: the initial focus upon the curriculum concerned, followed by probing outwards as necessary; and the joint responsibility of evaluator and evaluated for on-going assessment and reassessment of course aims. The results of applying this approach have been discussed above, but one further result in the case of chapter 6 was the use of specific and focussed evaluation methods such as study diaries and analysis of data by an evaluation panel. Methods such as these had been absent from evaluations of individual study courses until now.

7.3 General Conclusions

Both in the area of curriculum innovation and of educational evaluation, the results summarised above have much general relevance. To start with, an important point about teaching method innovation arises from the experience of manipulating elements of the Keller plan, reported above. It is clearly
inadequate to base plans for teaching methods upon theories from cognitive psychology about how individuals learn, without considering also the influential social setting in which undergraduate teaching takes place. However, it should be emphasised that the diversifications of the Keller plan for example, although based mainly upon pragmatism and intuition, have tended not to run counter to the Skinnerian theories which Keller held as so important. Hence the research has at various times emphasised consideration of social and cognitive psychology, and pragmatism in planning curriculum innovation. Further, the experiences that I have reported show that this lesson is particularly relevant to assessment planning, which in itself has powerful effects upon students and their studying.

The research has also shown that the process of curriculum development and innovation is rarely as methodical and predefined as the early protagonists of systematic curriculum development (e.g. Tyler (1949)) might suggest. Both teachers' aims and methods changed during the implementation of individual study courses, partly as a result of evaluation. This is in line with Rowntree's notion (1976) that evaluation should be at the root of curriculum development, and shows the likely consequences of such a strategy. The point about changes in teachers' aims is important for another reason also, for if the pattern described above should also be applicable to individual study courses in the U.S.A., this would call into question the value of many current psychometric evaluations, including those surveyed by, for example, Kulik (1976). Furthermore, the problem may be wider than just individual study courses, since it seems likely that changes in teachers' aims (perhaps similar to those described in this research) may take place as teachers adopt almost any kind of teaching innovation.

The findings of the workload study also have general relevance. The research suggests that the pattern of student study differs markedly between innovatory and non-innovatory teaching methods. This being so, links between features of teaching method and the study patterns that result, and links
between different study patterns, different students and
different subject matter may prove important in generating a
theory of independent learning, and a way of matching the
method to the subject and the student.

Perhaps predictably, it is more difficult to point to
tangible results of general relevance in connection with evalua-
tion. One point which originated in the dimensional analysis
of evaluation and continued throughout was the importance of
seeing evaluation as a continuum of possible approaches rather
than two or more opposing camps.

Formative evaluation prompts teacher and student aware-
ness, and as a result courses change considerably as they become
established. This pattern of change means that in general,
early summative evaluation of a teaching innovation may be
inappropriate, for the picture that results will be of a course
no longer in existence, no matter how accurate the picture is.

It is possible to draw out from the work reported some
lessons about substantiating this kind of research. In particular,
the test/tutorial study showed the wide range of evaluations of
the same data made by the panel of tutors, and it seems likely
that similar variations might appear in many other areas. Thus
the future of substantiating evaluations may depend on adopting
this kind of wide ranging data interpretation. Indeed I suggested
earlier that contrasting interpretations of this kind may be the
best information the evaluator can provide for teachers who want
to make a summative decision as to whether to try out an innovatory
teaching method. However, the purpose of most of the evaluations
has been formative, for I argued above that a summative "portrayal"
of individual study courses would have been of limited value.
This being so, the important test of the evaluations relies less
upon providing a well substantiated and accurate portrayal of the
method, and more upon the influence of the evaluation upon develop-
ing the method as far as possible.

Lastly, I have shown that practising teachers can engage in
worthwhile course evaluation, and that this process has both
strengths and weaknesses. Thus it is of general relevance to
suggest that future curriculum projects and initiatives in teaching
innovation, give at least part of the job of evaluation back to
the teacher himself.

7.4 Further Outlook

This last section examines the possibilities for future
research arising from the research I have reported (see 7.2 and
7.3), and draws the thesis to a close.

There are several possibilities for further work on individual
study courses. In the not too distant future it would be appro­
priate to survey the use of individual study courses once again,
for numbers have grown, diversification has continued, and new
subject areas (including the humanities) have adopted the method.
The possibilities for this are currently being looked into at
Surrey University.

The tentative studies of workload and tutorial contact on
individual study courses suggest other extensions. Attempts to
link work pattern with appropriate subjects, types of student,
and desired outcomes which were mentioned earlier may be parti­
cularly rewarding. Results might include a clearer picture of
the essential features of individual study courses, and the
formulation of theories about individualised instruction. While
it is important to extend the, at present rather tentative
tutorial study, the fact has to be faced that this might merely
confirm the picture of variation and subjectivity in the original
results.

In retrospect, the Keller plan and to a lesser extent the
audio-tutorial scheme have provided a valuable and relatively
safe starting point for curriculum innovation. But as experience
grows of modifications to such methods which suit the academic
scene in the U.K., more and more diversification may take place.

Possibilities for future evaluation also emerge, including
the need for a summative evaluation of the achievements of the
entire curriculum development which I have worked with. Microevalua­
tion is only partly tested and will need further trials, as will
the possibility of some interplay between Microevaluation and
larger scale evaluation. More generally, this thesis has been concerned with large and small scale evaluation by teachers and outside researchers, and it has discussed several other dimensions of evaluation. An important task for the future will therefore be to accumulate experience about when each type of evaluation is appropriate. Lastly, teacher evaluations and approaches such as Monitorkit merit considerable attention in future. This has just started in the fields of university lecturing and learning resources centres (Elton (1977)), but the possibilities range further, into training and education at many levels.

Drawing innovation and evaluation together points to an idea which may be important for the future of individual study courses in the U.K. where, in the context of widespread educational economies, they seem unlikely to produce savings over the lecture course. Essentially, individual study courses may have to be seen as a teaching method in their own right, achieving goals different from those of lectures rather than competing with them. For many of the goals described in the thesis are vital to the training of a scientist or an engineer. The parallel with undergraduate laboratory projects is inescapable. They too are certainly not more economical than scripted experiments, which predominate in the laboratory, and they are also justified because they manifestly involve worthwhile processes, although their outcomes are long-term and difficult to test. The future of individual study courses may depend on their being given a similar position in the science curriculum.

The research started off with a dual focus; innovation and evaluation. However, perhaps predictably the results emphasised integration: established and new forms of evaluation; an American innovation and a British context; outside research and inside monitoring; and learning theory and teaching practice.
Allen, P.S. (1976) "Reinforcement using the Keller Plan"  

Austin, S.M. and Gilbert, K.E. (1973) "Student Performance in a K.P.  
Course in Introductory Electricity and Magnetism."  
American Journal of Physics 41 12-18

Avey, H.P. and Northcott, P.H. (1975) "PSI Physics as a Service Course"  
The Australian Physicist July 1975 83-85.

Course in Introductory Psychology for University Freshmen".  
Int. J. Psychol. 1974, 11, (3), 183-201

N.F.E.R. Windsor.

Becker, H. et al (1968) "Making the Grade" New York, Wiley & Sons


Bijou, S.W. Morris, E.K. Parsons, J.A. "A PSI Course in Child Development  
with a Procedure for Reducing Student Procrastination".  

Black, P.J. (1976) "Aims, Processes and the Engineering of Teaching"  
Stud H. Ed. 1(2) 149-158

Black, P.J. (1977) "Individual Study in the Laboratory" in Bridge & Elton (1977)

Black, P.J. and Boud, D.J. (1977) "Counting the Costs" in Bridge & Elton (1977)

Teaching Methods Unit.


Bori, C.M. (1974) "Developments in Brazil" in Keller, F.S. and Green, B.A.  

Instructor Course". Pub : Centre to Improve Learning and Instruction,  
University of Utah, U.S.A.

Boud, D.J. (1973) "The Laboratory Aims Questionnaire - a New Method for  
Course Improvement". Higher Education, 2 (1) 81-94.

Courses" Surrey University.

Boud, D.J. (1976) "Investigation of Course Evaluation Methods in Higher  

Boud, D.J. and Bridge, W.A. (1974) "Keller Plan : A Case Study in  
Individualized Learning" in Nuffield Foundation Resource Package :  

Boud, D.J. Bridge, W.A. Maas, P. and Stace, B.C. (1975) "A Potentiality  
of the Keller Plan - The Transfer of Courses" Physics Education, 10,  
380-382.


Cassanova, J. et al (1977) "Report about the 'Keller Plan' in Teaching Thermodynamics". Translation of Report to the Faculty of Sciences; Dept. of Fundamental Physics, University of Valladolid, Spain.


Pub. by: Meredith Corporation, New York. U.S.A.

"Long Term Effects of P.I. in an Introductory Psychology Course".
Paper at 41st Annual Meeting of Eastern Psychological Association,
New Jersey, U.S.A.

Cronbach, L.J. (1965) "Evaluation for Course Improvement"
Teachers College Record, 64, 672-683.


D'Arruda, J. (1972) "A Self-Paced Program of Instruction at a Two Year
Institution". Internal Paper. University of Wisconsin,
Centre-Richland, U.S.A.

Davis, E.R. (1976) "The Role of Self-Paced Self-Study in Undergraduate


Davies, I.K. and Hartley, J. (1972) "Contributions to an Educational

Dowdeswell, W.H. (1972) "A Curriculum Development Project at University


Published by Centre for Advanced Study of Educational Administration,
Eugene, Oregon, U.S.A.


Elliott, J. and Adelman, C. (1973) "Towards a General Methodology of
Inquiry/Discovery Teaching". Ford Teaching Document I. CARE,
University of East Anglia, Nov. 73.*

London.

Elton, L.R.B. (1973) "Self Instructional Methods in University Science
Teaching". Paper presented at S.R.H.E. European Research Seminar,

Elton, L.R.B. (1976) "The Future of Educational Technology"
Internal Seminar, I.E.T. University of Surrey.

Elton, L.R.B. (1977) "Educational Technology: Today and Tomorrow".

Elton, L.R.B. (1977) "Handout on Self-Evaluation of Teaching"
Private Communication.

Elton, L.R.B., Boud, D.J., Nuttall, J., and Stace, B.C. (1973)
"Teach Yourself Paradigm: The Keller Plan". Chemistry in Britain,
9, (4), 164-168.

Embling, J.H. (1976) "The Schools and Educational Technology" in
Howe & Romiszowski (1976).


Hollinshead, B. (Ed.) (1977) "Independent Learning: Variations on a Theme" Occasional Paper 2, Staff Development Unit, Manchester Polytechnic.


Laszio, P. (1976) "What is the Keller Plan?" Internal Paper. Chemistry Institute, Liege University.


Micah Wei - Ming Leo, (1973) "Chemistry Teaching by the K.P." Journal of Chemical Education, 50 (1). 49.


Motazed, B. (1971) "A Learning System for Applied Statistics"
M.Sc. Thesis at University of Texas.


O'Connell, S. Penton, S.J. and Boud, D.J. (1977) "A Rationally Designed Self-Service Laboratory Mini-Course" Programmed Learning 
& Educational Technology, 14 (2) 144-161.

Ogborn, J. (Ed.) (1977) "Practical Work in Undergraduate Science"


Pennington, P. (1972) "Some Steps Toward More Flexible Instruction"


Prentice-Hall, New Jersey, U.S.A.


Ramsden, P. (1976) "Students Heed Their Teachers' Advice" THES 30.1.76.


Scriven, M. (1972) "Pros and Cons about Goal-Free Evaluation"


Times Higher Educational Supplement (1977) "Pre-Testing at the O.U." 6.5.77.


Traveggia, T.C. and Hedley, R.A. "Teaching Really Matters, Or Does It?" Engineering Education, 62 (6), 546-549.


ACKNOWLEDGEMENTS

The work reported in this thesis was supported by the Higher Education Learning Project in Physics, and benefited greatly from the resulting co-operation. I am also grateful to present and former members of the Institute for Educational Technology of the University of Surrey who have contributed their support, and companionship. The responsibility for the research however is mine alone, although in the case of chapter 5 section 4, the early work owed much to collaborations with D. M. Laurillard of the Institute for Educational Technology. I also acknowledge D. J. Boud and L. Willoughby whose work provided a worthwhile and valuable background to the research; and lastly, my supervisor, Professor L.R.B. Elton who made so much of it possible.
APPENDICES

I  Sample Materials

I/1 Individual Study Courses Survey Questionnaire (1975)

I/2 General Course Questionnaire: Particle Mechanics (1974)

I/3 Student Workload Diary (1976)

I/4 Sample Transcript from Keller Plan Tutorial Study

II  Monitorkit (second edition): Bound into back cover
This questionnaire is intended for any teacher in Higher Education who uses a course, or courses, in which the students cover the bulk of the material by individual study from printed materials, instead of attending a lecture course.

The purpose of this survey is to determine the main characteristics of self-study courses being used in Higher Education in the U.K., and draw together the combined experiences of those concerned.

If you run more than one self-study course, please complete a separate questionnaire for each. I can supply further copies.

If there are other teachers using self-study teaching methods in your institution or area, please note their name(s) on the back of this sheet.

You will be sent a report of the survey when it is completed.

Individual replies will be strictly confidential to myself.

Please return the completed questionnaire in the enclosed Freepost envelope to:

Mr. Will Bridge,
Freepost,
I.E.T.,
University of Surrey,
Guildford, Surrey GU2 5XH

Tel. Guildford 71281 ext. 881
1. **BACKGROUND**

1.1 Your Name .............................................

Department and Institution .................................................................

1.2 What is the subject of your self-study course? ..............................

Please circle its level: 1st year 2nd year 3rd year

Other, please state: .................................................................

How long have you used self-study teaching methods (circle as appropriate)

less than 1 year 1 - 2 years over 2 years

1.3 How many students study this course (please list any important year variations)

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

Please state the subject and qualification they are studying for:

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

1.4 Please add any further relevant background material (e.g. if they are part-time students, or the subject matter has changed, etc.)


2. **AIMS**

2.1 Before you started, what were your reasons for trying out a self-study approach in teaching this course?
2.3 Now that you have had some experience of running such a course, what aims do you now think it achieves?

2.4 What aims does it not achieve? In what areas were you disappointed?
3.1 Did you previously run this course as a lecture course? YES NO
Other, please state:

3.2 How is the course time-tabled in terms of length and number of periods per week?

3.3 Please give details of any other staff or postgraduate help for the course:

3.4 Please outline the number of units (or modules etc.) in the course:

3.5 Are there any further organisational details to add?

----------------------------------------------------------------------------------------------------------------------------------------

THE CHARACTERISTICS OF THE COURSE

If you feel your course is more or less based on the Keller Plan (or PSI), please complete section 4, and omit 5. If it owes little or nothing to the Keller Plan, omit 4 and complete 5.

----------------------------------------------------------------------------------------------------------------------------------------

4. BASICALLY KELLER PLAN

4.1 What made you decide to run this course by a Keller Plan approach?

4.2 I have listed over the five main characteristics of the Keller Plan in order to get a picture of your course. Please note briefly which characteristics exist unaltered in your course, and which you have changed or developed (and how). Note also any characteristics your course does not contain at all.

As far as possible, please give reasons for any omissions and change from these characteristics.
4.2.1 An important goal of the course is that the students should really learn (master) each part before they move on. To do this, they are tested before they are allowed to proceed, and they must get almost all of the test right if they are to pass.

(Please state changes from this with reasons)

4.2.2 The students can go through the units at any speed they wish, no pressure is brought to bear upon them to work in a particular way. (state changes with reasons)

4.2.3 The few lectures which are given are not essential parts of the course, and are regarded as a reward. (Please note if you do not lecture at all, and changes with reasons).

4.2.4 All of the content of the course is in the form of printed materials such as written units and books. (State changes with reasons).
4.3 Please state any other differences between your course and the original Keller Plan ideas as you understand them.

4.4 Please indicate (by ticking the relevant box) the frequency with which the following components have been used in your course units (sometimes called 'study guides'). In the comments section, please summarise your attitudes to each component.

<table>
<thead>
<tr>
<th>Component</th>
<th>Always used</th>
<th>Sometimes used</th>
<th>Never used</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Preknowledge for the unit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List of Aims or Objectives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An Introduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Procedure List</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Test Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solutions to Self-Test Questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments: pto
COMPLETE EITHER SECTION 4 OR SECTION 5. NOT BOTH.

5. SELF-STUDY COURSES NOT BASED ON THE KELLER PLAN

5.1 Were you aware of the Keller Plan approach when you started the self-study course?

YES  NO

5.2 If YES, why did you decide not to use it?

5.3 Please briefly describe the materials you provide the students with to help and guide their learning. Do they contain any of/all of the components listed in 4.4?

5.4 Please briefly describe what happens during the scheduled class meetings.
5.5 Please state whether there are any other features in the course, such as lectures, laboratory exercises, etc.

5.6 Do you check that the students have learnt the material as they proceed? YES NO

If YES, please say how you check this:

5.7 What determines the rate at which the students move through the course? Is there freedom of pace?

---

PLEASE COMPLETE SECTIONS 6 AND 7

6. REVIEW

6.1 From your experience of using a self-study course (of any type), which aspects of it do you consider to be vital if it is to succeed, and which are marginal?
6.3 If you have made any important changes to the course since you started running it, please explain them and say why you made them and what effect they have had.

7. **ASSESSMENTS**

7.1 How are the students assessed on this course?

7.2 Have you taken any steps to formally monitor this course, and evaluate it?

If YES, please outline what you did:
(a) Student Examination Results

(b) Their attitude towards the course

(c) The Attitudes of other staff members towards this course

(d) Any other areas in which you have noticed effects

Please add below overleaf any further details of the course you think would help me, your future plans for self-study courses, and any other comments.

Thank you very much,

W. A. Bridge
Please rate the following by circling or ticking the appropriate response. In each case briefly state reasons for your answer.

1. The Units

   In order to find out clearly how the course was to be run, roughly how many times did you have to read unit 0 at first?

   just once  2-3 times  4-5 times  more

Why? What were the difficulties?

Please rate on a five point scale:

<table>
<thead>
<tr>
<th></th>
<th>very useful</th>
<th>no use at all</th>
<th>REASONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>The list of objectives for each unit</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The notes in the units</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The problems in the units</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On average what proportion of the total number of problems in a unit did you try?

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

The book by Elton       1 2 3 4 5

Please comment on this and name and rate any other book you have used.

The film 'frames of reference' 1 2 3 4 5
Please rate the laboratory exercise 2a:

Please rate the computing exercise:

Please rate the laboratory exercise 2a:

Please rate the computing exercise:

Please comment on the difficulties of each:

Do you feel that this Keller Plan course helps you develop efficient study methods?

very much 1  2  3  4  5 very little

Please describe these methods, and suggest whether they also help you on other courses.
When studying the book references for a unit do you make notes:

always   sometimes   rarely   never

2. Tests and Test Sessions

Do any of the following put you off attending test periods and/or taking tests: anxiety about taking a test; the failing of a test; queuing; other factors? Please explain your answer:

Have you ever decided not to come into class because of the queues which sometimes form?  

YES  NO
to

P P P
to P rf <
P p CD
o
p

rf &q

3
o

CD

O

H-

H s; <
3 H* CD
P rf •-t

O

-<

H, O

Hj P
to H* P
O

O
O
p

H- rf
CD HP

H rf O
to
O X H*
in
P

O

G

1 -3

CD to

H P

to o to
to p
rf rf p

P

h »v-j.

P P Q
rf <n

X

to 3

H -

ri- 3 ^
P rf H*

rf

o
rf

H*

G o
P p

P

to

G

rf
P 3

3
O

O

3

^0 4

G CO O
•o rf

H P

I—
13
C
D P
O rf
rf rf
p O
P P
O
P
O Q
O rf
P rf
P G
to P
O
•-o rf
p
P
P

*-<
O
P
<
—
P
a.
o
P

to

rf
O
O
Pu

that:

P
o rf
P
p O
p c+
£u
H* •

H-

think

£0XW'llT4-'IOM

p

O
O

rf

you

p

— > rf P
rf t
d
p
P o
rf o CD o G
f—
•
O rf
H r-* in o c+
H- H rf to CD
P O P o to
!—
> to H> p rf
CO
H>
to
•P HH
o
H-i r f o P

Do

t-3
P1
t_ Hr
O to

CD
’P

rf
P
G

H*

to
P
P
P t_s.
G
O
rf
P

f—1*3

a> p
to o
to p

ip

ip

CD O

tp

ip

to s;

to p

p
rf

P
P
to O

OJ

03

p

03

rf

03

rf
P

O

4^.

O
H

CD

CD

tn

cn

03

tn

03

3 3
O P
H O
G P

n

oe*
•
tp
p
-3
03


The staff tutors 1 2 3 4 5

The Post Grad. tutors 1 2 3 4 5

The Undergrad. tutors 1 2 3 4 5

In terms of the standard they require for you to pass a test:

very 

high

very 

low

The staff 1 2 3 4 5

The Post Grad. tutors 1 2 3 4 5

The Undergrad. tutors 1 2 3 4 5

How confident are you that you really know the subject matter of a unit once you have passed the test?

very 

confident

fairly 

confident

unsure

not at all 

confident

If you are not very confident, what could be done to improve this?

Please state how many times you have acted as an undergraduate tutor

N/A 0 1 2 3 4 5 6 7 8

If applicable, why did you volunteer, and what have you gained from tutoring?
3. **Workload**

Compared with the average for other courses which you take, does this course require:

- much more
- 1 2 3 4 5
- much less
- of your time
- of your time

How much time did you spend on this course over the past week, not including class time?

- less than
- 1 - 3
- 3 - 5
- more than
- 1 hour
- hours
- hours
- 5 hours

Was this more or less than a typical week?

- much more
- more
- same
- less
- much less

If your test passing rate or attendance at class is starting to fall, can you suggest why?
The laboratory exercise?..........................
The computer exercise?..........................
The term paper?.................................

How many times have you attended (a) The film 'frames of reference'? ..... (b) The optional lectures? \\

How often do you attend test session on average?

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 per week</td>
</tr>
<tr>
<td>1 per week</td>
</tr>
<tr>
<td>1 per 2 weeks</td>
</tr>
<tr>
<td>less than 1 per 2 weeks</td>
</tr>
</tbody>
</table>

Please comment as to how the system of marks on this course influences your study, are there any aspects of it which seem unfair?

4. Organisation

What advantages and disadvantages have you found in the fact that you can choose the subject and order of the units you study?
How do the deadlines influence your work on the course? If you have ever fallen behind, please say how this affected you.

Do these deadlines provide sufficient compulsion in the course for you?

- too
- slightly
- about
- slightly
- too
- much
- too much
- right
- too little
- little

Next year, if a particular course were available taught both by Keller and Lecture Method, would you choose to be taught by Keller Plan?

YES  NO

Please comment on this and use the reverse side of this sheet for general comments on the course.

Thank you for your co-operation and your time,

Will Bridge
Study Diary

Name : ......................................... (This will be confidential to the researcher)

Week : .........................................

Read these notes please.

1. Fill in this diary each day for 7 days, including weekend.

2. Fill in your lecture/lab/tutorial timetable for the week in the column marked "ACTIVITY" now. Indicate that this is scheduled classtime by underlining it. E.G. in the Monday, 10-11 a.m. "ACTIVITY" column might be - Maths Lecture. If you did not attend it though, or it only lasted half-an-hour, note this in the space on the page below the table.

3. When you are doing nothing to do with the courses (such as sleeping, shopping, etc) just put an X in the "ACTIVITY" column.

4. For times when you did individual study, i.e. "homework", write the word "STUDIED" in the "ACTIVITY" column followed by the subject studied. Write "studied - general" if you were just, say, reading a journal on physics or engineering.

5. I am particularly interested in the times when you do INDIVIDUAL STUDY. Please put full details in the second column, giving exact times, and note exactly what you do.
For example:

Example of individual study time notes

8 - 9 p.m. Studied Maths Spent from 8.15 to 9.15 reading the maths lecture notes then went on to start chemistry lab. report until 9.45.

9 - 10 p.m. Studied Maths Then chem.

6. To repeat, fill in this diary each day (perhaps best last thing at night) and be as full as you can in your description of individual study time.

7. Return the diary to me this time next week. To say thank you for doing a good job, I will pay £1.00

W. A. Bridge
<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>FULL DETAILS OF INDIVIDUAL STUDY TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 - 9 a.m.</td>
<td></td>
</tr>
<tr>
<td>9 - 10 a.m.</td>
<td>Lecture Period 1</td>
</tr>
<tr>
<td>10 - 11 a.m.</td>
<td>Lecture Period 2</td>
</tr>
<tr>
<td>11 - 12 noon</td>
<td>Lecture Period 3</td>
</tr>
<tr>
<td>12 - 1 p.m.</td>
<td>Lecture Period 4</td>
</tr>
<tr>
<td>1 - 2 p.m.</td>
<td>Lecture Period 5</td>
</tr>
<tr>
<td>2 - 3 p.m.</td>
<td>Lecture Period 6</td>
</tr>
<tr>
<td>3 - 4 p.m.</td>
<td>Lecture Period 7</td>
</tr>
<tr>
<td>4 - 5 p.m.</td>
<td>Lecture Period 8</td>
</tr>
<tr>
<td>5 - 6 p.m.</td>
<td>Lecture Period 9</td>
</tr>
<tr>
<td>6 - 7 p.m.</td>
<td></td>
</tr>
<tr>
<td>7 - 8 p.m.</td>
<td></td>
</tr>
<tr>
<td>8 - 9 p.m.</td>
<td></td>
</tr>
<tr>
<td>9 - 10 p.m.</td>
<td></td>
</tr>
<tr>
<td>10 - 11 p.m.</td>
<td></td>
</tr>
<tr>
<td>11 - 12 midnight</td>
<td></td>
</tr>
</tbody>
</table>
Student comes up to tutor's desk, shows him the test form.

Can you mark this one?

Yes. Sit down.

Here you are then.

(tutor looks through the answers. Comes across a question ....)

Well you don't really know the answer to that one because you don't know whereabouts it is between the two bodies. Well you don't really know which other force is acting on this one do you? What are they? Write them down.

Um, this you do don't you?

Well, write it down.

You've got your moon and you've got your earth, and you've got a rocket.

right, right. So how's that being affected? How is the rocket being affected by these two?

Well, you've got a force of attraction between that and that; and that and that.

So it's being accelerated towards?

Here at this point, and after that towards that by gravitational attraction. So .. (pause) .. it wouldn't be going at constant velocity.

Correct.

Cos at this point it would accelerate towards the moon.

Right, that's absolutely it. So in general it's experiencing gravitation field of attraction to both of these at one point in space. Right, ok. Now let's have a look at the next one. That's correct .. (pause) yep, o.k. (to the next one). Did you remember that or did you work it out?

I worked it out from the diagram.

Ah yes, you worked it out at the bottom, didn't you.

That was my first inertial frame.

I see, why didn't you put that (and he drew it). So would you like to explain that to me?

Well if this is the velocity relative to this inertial frame ...

You're calling that that, are you?

Yes. And that's the velocity of the second inertial frame.

With respect to that .. wait a minute..

With respect to that - yes.

Right.
Yeah well why do you draw that particle over there? It's the same particle isn't it?

Yes, but it helps to visualise it.

Oh no, it only helps to confuse it, doesn't it? That particle has two velocities, depending on the frame in which you measure it.

Ah well, in that case it doesn't have the two diagrams does it not?

No, one diagram but two velocities. You label that as you usually do, \(-A\) - and that as you usually do - \(A'\) - then surely that's what the two velocities are of the particle.

Well what symbol do you use?

Oh \(u\) primed.

Yes that's right. I think you're only confusing the issue by drawing the separate particle with respect to that frame of reference. After all there is only one particle.

Yeah.

So it's the velocity of that particle in a given frame of reference.

Mmmmm.

And in the second frame of reference the velocity of the particle relative to it, what is the velocity of that relative to it? What is it (pause) \(u\) primed.

(agreeing) uh-uh.

And then you've got to establish a relationship between \(u\) primed, \(u\), and \(v\).

Well it's going ... going that way and it's obviously the difference between the two. (student laughs)

Obviously.

(pause)

let's call the particle \(p\).

I see, you draw that relative to that, do you? Right, I'll do that then. Here goes. The velocity of um um ... that relative to that, that say is the velocity of \(o\) primed relative to that.

Yeeees. Well I think you'd be better if you turned it the other way round, otherwise you might later get problems with 'sines. Cos that velocity is the negative of that velocity, isn't that so? I mean, you can have it that way round if you like. What do you want to do?

It's alright. I'll turn it round.
Then you've got the velocity of the particle (he draws it, pause) and then the velocity of that frame, then you've got the velocity of the particle \( p \) relative to that ...

Right. Good.

... is that. (and he writes something)

And now you've got a third thing to specify ...

Yes, you have. It's the velocity of this particle relative to your original whatchamecallit.

Right.

So obviously the sum of these is going to be that.

Mm. Well I think that the way you've done it you might have a sine problem.

Alright then.

Yes. Can you draw up a diagram showing that now?

Right then. Now what do we want? We want the velocity of this thing along here plus the velocity of that particle ....

Well mark that with one of these symbols. Mark that.

With one of these symbols?

No, with one of these symbols. You know, \( u \) or \( v \), \( u \) primed or \( v \) primed.

Like this?

Yes. That's correct.

\[ -o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o \]

So you have that \( v \) equals \( u \) plus \( v \) primed.

Yes.

Which is what you've got there. You rearrange that because you are asked to express it in terms of \( u \) primed, which gives you that, yes? Alright?

Yes.

(they pass on to the next question, which the tutor ticks off silently without querying. Move to the third question and the tutor looks at it.)

Ah-ha. That looks quite good. (looks at the diagrams the student has drawn) You've got two points, \( p_1 \) and \( p_2 \), and you've got these two which are called \( r_1 \) and \( r_2 \), and you've got to establish the distance between them.

And the distance between them is given by that equation there.
Correct. Now if you go on to here, in this frame of reference, then $v_2l$ equals (he goes through checking his algebra), yep. (then he goes onto next bit). That's what you're calling $r$ primed is it?

es, that's right. Now in another frame then $r_l$ primed equals $r_l$ minus $vt$, and $r_2$ primed equals $r_2$ minus $vt$. Correct. So the difference between the points is now $r_2$ primed minus $r_l$ primed. (the tutor continues) then what's that? Oh I see what that is. It's just a spot off the paper isn't it? (tutor goes on) Equals $r_2$ primed minus $vt$ minus $r_l$ plus $vt$ - those two cancel out. That prime shouldn't be there.

You're a bit slap happy with your primes, aren't you? (pause) and therefore you've proved what's required haven't you?

That's good. (looks back over the whole paper)

How've you done? Well, you got that one right, and you got that one right, after a bit of explanation. Well, you weren't in fact asked to explain it, so that you got that one right as well. (pause) Good. (pause) Do you have any problems with this part of the course.

ah.

Then. Good.

Student prepares to leave)

wait a minute. I ought to write your mark down on the gray sheet as 1. (he does so, remarking that the student is only two days in hand of the deadline). Right, thank you, goodbye.
A Resource Package for University Teachers who wish to Monitor their own Courses, particularly Self-Study Courses. Produced in conjunction with the Higher Education Learning Project in Physics - H.E.L.P. (P).

CONTENTS

PART I COURSE MONITORING - AN INTRODUCTION

PART II IDEAS AND EXAMPLES

QUESTIONNAIRES
1. Feedback Slips
2. General Course Questionnaires
3. Essays on the Course

DISCUSSION/INTERVIEWS
4. Teacher and Student
5. Teacher and Outside Adviser
6. Teacher and Groups of Students

RECORDS TAKEN BY THE TEACHER
7. Student Progress Data
8. Record Keeping

PART III A STRATEGY FOR ASKING QUESTIONS
WHY?

Teaching is said to be a performing art. However, unlike theatre or opera, the audience doesn't clap, the critics don't often slate or acclaim, and the attendance figures don't mean much. Feedback about teaching can be found in tutorials, question sessions, and examinations, but the message is not always strong or convincing. Thus, the simplest reason for using course monitoring is to encourage a systematic feedback of information about the dynamic process of teaching.

WHAT?

Monitorkit is made up of a collection of materials intended to help teachers monitor their own courses. It consists of sample materials and ideas for possible methods and overall strategies. It also gives some indication as to their likely outcomes.

I have used the term "course monitoring" rather than "evaluation" to emphasise that this package does not attempt to promote time-consuming or esoteric research activities on the part of the teacher. Rather it is an essentially practical approach to the problem of providing teachers with more systematic information about their teaching, and their students. This problem is particularly important in the case of innovative, and unfamiliar teaching methods such as the Keller Plan, or other forms of self-study. Indeed this package is primarily intended for use on such courses. However, experience has shown that much of the material is more widely useful, and it seems likely that traditional lecture courses, for example, would benefit from the use of certain parts of the package.

Like much of the H.E.L.P. (P) material, Monitorkit is designed to generate questions and ideas rather than provide complete answers. The package is intended to be both read, and used, selectively, with only a fraction of the ideas, examples and procedures being used at any one time. Thus on first reading through Monitorkit, I suggest that you read only Parts I and III in full, and select just a few of the sections in Part II to read. Part III is particularly important since it suggests how different monitoring methods can be built up into an overall strategy.

It is important to realise the kinds of information that Monitorkit can provide. Perhaps the majority of measures outlined provide, in a pre-planned and systematic way, subjective data from students and tutors, such as their likes and dislikes, difficulties and reactions to the course. A minority of the measures provide more objective data about student progress, achievement, and learning.

In Part II a distinction between two categories of subjective data is
drawn, and this is more fully explained in Part III. At this point it is adequate to say that FORMATIVE data is used to make on-going improvements in a course, whilst SUMMATIVE data is used to form an overall view of the course, and help judge its overall worth.

HOW?

Having decided that course monitoring is important for a particular course, it is actually carried out in five stages:

1. Deciding upon the aims of the proposed monitoring procedure.
2. Devising a programme to achieve these aims.
3. Putting the programme into practice.
4. Analysing the results.
5. Taking action on the basis of these results.

Part II, which constitutes the bulk of Monitorkit, is devoted to items 3 and 4, while Part III aims to cover items 1, 2, and 5. Thus, I suggest that you read just a few sections from Part II, and then move on to Part III.

BACK TO ME

This second version of Monitorkit remains a very tentative step into the field of course monitoring. To do better, I need your feedback about it, including all-out criticism! I hope you will keep me in touch with your work, and let me know if I can be of any further help.

Good luck - I hope you find some useful ideas here.

Will Bridge,
Institute for Educational Technology,
University of Surrey,
Guilford, Surrey,
GU2 5XH

N.B. This material is still in the development stage, and results from it may not be used outside the relevant course without permission.
PART 2  COURSE MONITORING - IDEAS AND EXAMPLES

CONTENTS:

QUESTIONNAIRES

1. Feedback Slips
2. General Course Questionnaires
3. Essays on the Course

DISCUSSION/INTERVIEWS

4. Teacher and Student
5. Teacher and Outside Adviser
6. Teacher and Groups of Students

RECORDS TAKEN BY THE TEACHER

7. Student Progress Data
8. Record Keeping

1. FEEDBACK SLIPS

Contents:

1a Background Paper
1b Sample Coursework Unit Feedback Slip
1c Sample Laboratory Exercise Feedback Slip

* This technique is primarily of use in courses which have printed coursework units.*

1a Background Paper

Introduction

As you can see from the sample feedback slips in 1b and 1c, they are very short questionnaires covering a limited range of topics. The essence of feedback slips is that they are quick to complete, and therefore they can be used on several occasions during the course.

The information they provide is in the form of answers to detailed
questions; there is usually little opportunity or time for the students
to raise ideas or issues of their own.

Procedures

A range of possible uses can be listed:

1. They can be attached to the back of a particular coursework unit, (see example in 1b), for example, one which is newly written or troublesome. Alternatively they could be attached to all the coursework units the first time these are used.

2. Feedback Sheets of a different format can be issued with laboratory or computing exercises on the course (see example in Ic), to be completed in a spare moment whilst the exercise is in progress.

3. They can be issued at a particular meeting of the class. (Note that if the course allows free self-pacing, all the students do not complete (1) or (2) at the same time.) An example of the type of information that this would produce is the amount of time spent by students on the course during a particular week. In fact this technique can be used to gauge the student's reactions to problems as they emerge, or to aspects of the course, such as optional lectures, which occur at a fixed time.

In whichever of the three contexts you use feedback slips, the information which you get out of them will be of a detailed nature on specific topics suitable for improving the course as it proceeds. It is less likely to provide an overall perspective upon which the value of the whole course can be judged.

Construction Notes on how to go about writing both Feedback Slips and questionnaires appear in Part III Section 5. This presents more detail than is required for producing a short feedback slip where, in particular, the range of possible question formats is limited. In this case, it is probably wise to rely on a high proportion of numerical responses plus just one or two more open questions.

The problem of anonymity of feedback slips is not great. Generally speaking, the specific and detailed nature of the questions means that the students respond freely even when they are asked to sign the slips. If students have to give their names, the distribution and collection of feedback slips is made easier, and it is possible that the analysis will benefit from knowing the source of the responses. Also, it enables you to chase up missing slips, which is important, since you should try to obtain responses from all of the students. It is generally those most in need of help who fail to return questionnaires or slips.

Administration Notes on how to go about writing both Feedback Slips and questionnaires appear in Part III Section 5. This presents more detail than is required for producing a short feedback slip where, in particular, the range of possible question formats is limited. In this case, it is probably wise to rely on a high proportion of numerical responses plus just one or two more open questions.

The problem of anonymity of feedback slips is not great. Generally speaking, the specific and detailed nature of the questions means that the students respond freely even when they are asked to sign the slips. If students have to give their names, the distribution and collection of feedback slips is made easier, and it is possible that the analysis will benefit from knowing the source of the responses. Also, it enables you to chase up missing slips, which is important, since you should try to obtain responses from all of the students. It is generally those most in need of help who fail to return questionnaires or slips.

Analysing Feedback Slip Data

This is covered in Part III Section 6 but again to a greater depth than you may find necessary. In addition, since it is possible that a sequence of feedback slips could be used on one course, it is advisable to keep a running record of the results of each slip, and these results may be compared with the efforts you have made to improve the course on the basis of them. (see also 8a Sub-section 2).

Conclusion

The feedback slip is an easy, quick means of gathering formative data about the course as a whole, or parts of it, for the purposes of
improving it. It does not provide an overall view of the course. However, it is particularly well suited to locating problems and poor materials at an early stage.

Other sections of Monitorkit relevant to this method:
- Part III Sections 1 - 6;
- Part II Section 2b, and 8a subsection 2 (see diary/notebook suggestion in particular).
## Self-Study Course Feedback Sheet

**Name**

**Unit No.**

Please circle ratings, and add your own comments:

<table>
<thead>
<tr>
<th>Section No. in text books</th>
<th>Interest</th>
<th>Difficulty</th>
<th>Missing preknowledge</th>
<th>Comments on difficult problems etc.</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></td>
</tr>
<tr>
<td>1</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following Laboratory Experiments:

(i) The Wheel

(ii) The Pulley

Approximate time taken for this unit: 

............. hrs.

General comments on the text:

General comments on the problems:

*NB. Sample materials such as this are to be regarded as examples, not models.

(Based on an item first used by the Course Evaluation Research Project, I.E.T., University of Surrey.)
Feedback Sheet

Please rate the first three items on this questionnaire on the five point scales.

1. The difficulty of this experiment
   too easy 1 2 3 4 too difficult 5

2. Your interest in this experiment
   boring 1 2 3 4 stimulating 5

3. The relevance of this experiment to the chemistry courses is:
   unrelated 1 2 3 4 closely related 5

4. What knowledge or skills were you assumed to have before you started this experiment, that you did not have?

5. Please write down any particular difficulties which you had with this assignment:

6. Approximately how many hours did you spend in the laboratory doing this experiment?

7. Approximately how many hours did it take you to write up this experiment?

8. Any other comments?

*(this item designed by R. Watson, I.E.T., University of Surrey)*
2. GENERAL COURSE QUESTIONNAIRES

Contents:

2a Background Paper
2b Sample Evaluation Report
2c Sample General Course Questionnaire

2a Background Paper

Introduction

The type of two or three page questionnaire which is given to all the students at one time is a common way for teachers to monitor self-study courses. This paper outlines such questionnaires and suggests what they may be used for. It also gives advice on how to construct, administer, and analyse the data from general course questionnaires.

The general course questionnaire is usually given to the students on one occasion for them to fill in during their free time, and return at a later date. Section 2c gives an example of such a questionnaire prepared for a Keller Plan course. It may contain open-ended questions to which the student has to provide his own answer, or closed questions in which he simply has to choose a given response or rating, or a mixture of the two.

Uses

Five possible uses for general course questionnaires can be listed:

(1) Used in their most common way, they can either provide you with information about elements of the course which will help to improve it as it is running or provide a basis for judging the value of the course as a whole. A single questionnaire often performs both functions. To obtain information useful to the former purpose, the questionnaire questions should ask for criticisms and suggestions about particular aspects of the course. For the latter purpose, general questions about the whole course and its effect on the students should be asked.

(2) General Course Questionnaires can also be used as a starting point for a monitoring procedure which contains a number of elements, each building upon the information gathered in the questionnaire. It can be used to sift out the main issues which the students think are important in order to investigate them further. If questionnaires are to be put to this use, they should be open-ended, and allow the students to identify issues which they feel to be important.

(3) They can be used to detect the difference in viewpoint between the staff on the course and the students. This can be done simply by asking both staff and students to fill in the questionnaire. This may provide particularly useful information about aspects of the course which both staff and students take an active part in, for example, the Keller Plan test periods. Indeed, differences in viewpoint between staff and students, particularly on the understood aims of the course, often provide pointers to potential problem areas.
(4) They can be used to detect changes in outlook of the students as they go through the course. This is achieved by issuing similar questionnaires on two occasions, typically before the course, and when it finishes.

(5) They can be used to gather the views of particular students. There are two possibilities here. Firstly, they can be given to particular groups of students (typically those with difficulties in mathematics or language, or those following particular subject specialisms) in order to gauge the value of the course to the group. Secondly, a longer questionnaire on the subject of the whole class's reactions to the course may be given to a suitable individual student such as the class representative. He would be asked to gather all the students' views before answering the questionnaire. You might well help him to search out information, and gather together all possible viewpoints by providing him with previous evaluation reports on the method of self-study in use on the course. A sample report you might use appears in 2b.

Notes about questionnaire writing have been separated out from this section since they are of general applicability. They appear in Part III section 5. I hope they prove useful to anyone planning a questionnaire for the first time, and I don't propose to repeat them here. I feel that all the stages which appear in Part III section 5 should be followed in producing a general course questionnaire, although three points should be emphasised.

(i) The initial drafting of aims to be pursued in the questionnaire is particularly important for those longer questionnaires. Without it the quantity and diversity of information is hard to disentangle.

(ii) Deciding the question format is most important on general course questionnaires, and variety is probably worthwhile. No firm rules can be given, but I find the decision between open-ended and closed questions can be made as follows:

Broadly speaking, I find it adequate to ask a closed question (one in which the student has to fit his ideas into my words and alternatives such as a rating scale, response selection, or checklist) when I am sure that the alternatives I present, or the way the question is phrased represents and covers all possible student views or responses.

If I am in a little doubt about this, I add a supplementary open question after a closed one, but if I am in much doubt, I don't use closed questions at all, and I rely on open-ended ones.

For example, I would be satisfied with a closed question covering the number of times the student had been into the class period over the past fortnight.

I would ask a closed question supplemented by an open one to find out how many of the unit exercises the student had tackled. (He may have half-finished some, or only tackled part of each.)

I would just use open-ended questions to investigate how the students view of, say maths, had been changed by the course.
(iii) In this kind of questionnaire you are aiming to get wide-ranging comments and criticisms about your course, and implicitly, about yourself. This raises the question of confidentiality. Experience seems to suggest that by making such questionnaires anonymous, more information is obtained, and it is candid, and genuine, in nature. However, if the questionnaires are anonymous, it may be difficult, but not impossible, to make sure that you get a good response rate (see Part III section 5).

Analysis

Again the analysis of general course questionnaires has been separated out from this section, and appears in Part III section 6, as it is of more general use. Let me simply emphasise that it is particularly easy to put off the analysis and report writing stages when faced with a pile of questionnaires!

Conclusions

It is quite likely that you will decide to use a general course questionnaire, and, if you do, you should find the sections on construction and analysis (Part III sections 5 and 6), useful in this. Finally, remember the importance of an overall strategy. It will prescribe the questions you ask, and will avoid the "now, what would be a good question to ask?" syndrome, which often produces interesting but useless information (see Part III sections 1 and 2 for a discussion of monitoring strategies).

Other sections of Monitorkit providing useful information for this topic

Part III Sections 1 - 6
Part II Section 4b
OVERALL ATTITUDES TOWARDS THE COURSE

A range of attitudes from highly positive to highly negative about the Keller Course was quite evident. When the course was first introduced, this balance was overwhelmingly on the positive side. During the first term however, some students were put off by the amount of hard work they had to do, but it seems likely that the course's popularity was boosted at the end when students found revision easier than in other courses.

The most common reasons for positive feelings about the course were produced by the two freedoms - of content and pace, with the former being slightly more common. Other important factors which were mentioned were that the course was an efficient way of learning, and that it was good to know what was important and to be learnt. The incentives to steady working were also thought to be a major advantage of the course by many students. Again a final advantage was seen as the purposeful staff/student contact on the course.

Some students were indifferent to the course. They often made the comment that a Keller Plan course was much better than mediocre lectures. These students particularly were put off the course by the text book which they found to be dry and to contain certain poorly explained sections.

Students who disliked the course usually gave as the reason the time they had to spend studying on it. Slow students in particular found it took up twice as much time as conventional courses during term-time. Also students complained about the course organisation; they objected to queuing for tests and computer booking, and they also objected to the more blatant forms of cheating. In this context a particular phrase of complaint was heard extremely frequently - that the system of marks made the course seem childish and "like being back at school".

Over the two terms students on the course became very sociable and a group feeling seemed to develop. Students suggested two reasons for this - the nature of the course which allowed them to talk to their colleagues during scheduled periods, and the friendly informality of the staff on the course.

Foreign students had certain particular attitudes towards the course. They definitely found it easier to understand the content of the course by reading the units than they would have done by listening to lectures, and this was a large advantage to them. However, they sometimes found the test questions difficult to understand, particularly where careful interpretation of uncommon phrases was required to get to the point of the problem.

The staff and postgraduate tutors' attitudes towards the course varied considerably. The two senior staff members were clearly very committed to the method but certain other tutors were less sure (this came across to the students and emerged in interviews). The main advantage of the course as seen by tutors was the staff/student contact, and those tutoring for the second year felt that they could now give slow students the help they needed. The main disadvantage was seen as the complexity of the mark system, and the branching system, (probably because they had the job of implementing them), and being up-to-date with all the units.

The development and transfer of study habits by Keller Plan students has been considered. Some students do feel that they develop particular study habits in the Keller courses, and suggest increased self-discipline and the concentration on important or difficult areas as two main study habits which they develop. Asked about transfer of these study habits to traditional courses however, students remarked that they could not gauge this as the Keller course and traditional courses were as different as 'chalk and cheese'. On similar courses, (one example was a mathematics programmed learning course), a few students did feel that worthwhile study habits were transferred.
A KELLER PLAN COURSE

(You are welcome to copy and use any part of this questionnaire, but it should be regarded as an example, not a model.)

Name (optional) .................................................................

Department .................................................................

These questions are designed to find your opinions on certain aspects of the course and to give you a say in the way in which you are taught. If you cannot express yourself fully in the space provided for a particular question please continue your answer on the blank sheet opposite it. If a question does not apply to you, for example a question on a unit you have not studied at all, simply put N/A.

This questionnaire will be treated as confidential, and individual replies will not be disclosed to any of the academic or tutoring staff.

Finally please appreciate that the improvement of teaching is important to yourself, and that you can affect it. Please fill in this questionnaire thoughtfully and return it to me in the Particle Mechanics period before the end of term. Or place it in union p/h 'B'.

Thank you very much for your help.

W. A. Bridge,
Institute for Educational Technology
1. The Units

In order to find out clearly how the course was to be run, roughly how many times did you have to read unit 0 at first?

<table>
<thead>
<tr>
<th>just once</th>
<th>2-3 times</th>
<th>4-5 times</th>
<th>more</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why? What were the difficulties?

Please rate on a five point scale:

<table>
<thead>
<tr>
<th>very useful</th>
<th>no use at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

The list of objectives
for each unit

cc 2

The notes in the units

cc 3

The problems in the units

cc 4

On average what proportion of the total number of problems in a unit did you try?

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

The book by Elton

cc 5

Please comment on this and name and rate any other book you have used.

The film 'frames of reference'

cc 6
<table>
<thead>
<tr>
<th></th>
<th>very useful</th>
<th>no use at all</th>
<th>REASONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please rate the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>laboratory</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 7</td>
</tr>
<tr>
<td>exercise 2a:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please rate the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>computing</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 8</td>
</tr>
<tr>
<td>exercise:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>very enjoyable</td>
<td></td>
<td>not enjoyable</td>
<td></td>
</tr>
<tr>
<td>at all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please rate the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>laboratory</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 9</td>
</tr>
<tr>
<td>exercise 2a:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please rate the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>computing</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 10</td>
</tr>
<tr>
<td>exercise:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please comment on</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the difficulties of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>each:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you feel that this Keller Plan course helps you develop **efficient** study methods?

<table>
<thead>
<tr>
<th></th>
<th>very much</th>
<th>1 2 3 4 5</th>
<th>very little</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>very much</td>
<td>1 2 3 4 5</td>
<td>very little</td>
</tr>
</tbody>
</table>

Please describe these methods, and suggest whether they also help you on **other courses**.
Please state any units which you have found particularly good or poor, or any that you had particular difficulty with and state reasons for this.

When studying the book references for a unit do you make notes:

- always
- sometimes
- rarely
- never

2. Tests and Test Sessions

Do any of the following put you off attending test periods and/or taking tests: anxiety about taking a test; the failing of a test; queuing; other factors? Please explain your answer:

Have you ever decided not to come into class because of the queues which sometimes form?

- YES
- NO
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

This contact is more

In contrast, the staff have given you

Your time is more

The costs have been given

This contact is more

The subject is more

Your time is more

You understood the subject

Do you think that?
In terms of helping you learn during a test session please rate:

<table>
<thead>
<tr>
<th></th>
<th>very useful</th>
<th>no use at all</th>
<th>REASONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>The staff tutors</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 18</td>
</tr>
<tr>
<td>The Post Grad. tutors</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 20</td>
</tr>
<tr>
<td>The Undergrad. tutors</td>
<td>1 2 3 4 5</td>
<td></td>
<td>cc 21</td>
</tr>
</tbody>
</table>

In terms of the standard they require for you to pass a test:

<table>
<thead>
<tr>
<th></th>
<th>very high</th>
<th>very low</th>
</tr>
</thead>
<tbody>
<tr>
<td>The staff</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>The Post Grad. tutors</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>The Undergrad. tutors</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>

How confident are you that you really know the subject matter of a unit once you have passed the test?

<table>
<thead>
<tr>
<th></th>
<th>very confident</th>
<th>fairly confident</th>
<th>unsure</th>
<th>not at all confident</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If you are not very confident, what could be done to improve this?

Please state how many times you have acted as an undergraduate tutor

<table>
<thead>
<tr>
<th></th>
<th>0 1 2 3 4 5 6 7 8</th>
<th>cc 26</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If applicable, why did you volunteer, and what have you gained from tutoring?
Please make any general comments about the tests and/or test periods:

3. Workload

Compared with the average for other courses which you take, does this course require:

much more 1 2 3 4 5 much less
of your time of your time cc 27

How much time did you spend on this course over the past week, not including class time?

less than 1 - 3 3 - 5 more than
1 hour hours hours 5 hours cc 28

Was this more or less than a typical week?

much more more same less much less cc 29

If your test passing rate or attendance at class is starting to fall, can you suggest why?
Roughly how much time have you spent so far on:

The laboratory exercise? ....................
The computer exercise? ......................
The term paper? ..............................

How many times have you attended (a) The film 'frames of reference'? ......
(b) The optional lectures? ......

How often do you attend test session on average?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 per week</td>
<td>2</td>
</tr>
<tr>
<td>1 per week</td>
<td>1</td>
</tr>
<tr>
<td>1 per 2 weeks</td>
<td>1</td>
</tr>
<tr>
<td>less than 1 per 2 weeks</td>
<td>cc 30</td>
</tr>
</tbody>
</table>

Please comment as to how the system of marks on this course influences your study, are there any aspects of it which seem unfair?

4. Organisation

What advantages and disadvantages have you found in the fact that you can choose the subject and order of the units you study?
Have you ever started to study a unit and got stuck with it resulting in you going on to another.

| often | sometimes | rarely | never | cc 31 |

How do the deadlines influence your work on the course? If you have ever fallen behind, please say how this affected you.

Do these deadlines provide sufficient compulsion in the course for you?

| too much | slightly | about | slightly | too | cc 32 |
| too much | right | too little | little |

Next year, if a particular course were available taught both by Keller and Lecture Method, would you choose to be taught by Keller Plan?

| YES | NO | cc 33 |

Please comment on this and use the reverse side of this sheet for general comments on the course.

Thank you for your co-operation and your time,

Will Bridge
3. ESSAYS ON THE COURSE

Contents:

3a Background Paper
3b Sample Essay

3a Background Paper

Introduction

One way to get feedback from students about a self-study course, or indeed any course, is to ask them to write an essay or report about it. This scheme has two main advantages:

1. It does not require very much arranging on the part of the teacher.

2. The responses which it gathers will display the students' priorities and viewpoint, and they will not be influenced by the range of areas covered by, say, questionnaire questions.

There are three disadvantages to be overcome though:

1. Particularly if the essay is to be assessed, it will be extremely hard to be sure that the responses represent the students' true feelings, rather than their judgement of what is required of them.

2. Students find some difficulty in sharpening up their views about a particular course, and have difficulty in communicating their feelings unaided.

3. The feedback may be so diverse it is hard to systematise, or it may be difficult to draw any widely held conclusions from the information.

This method of course monitoring may tend to produce information of a predominantly overview type, about the course as a whole and its major elements. It is thus more useful in helping judge the success of the whole course, rather than improving parts of it. It is an attractive alternative to the general course questionnaire, and it may serve the same purpose. It would be particularly useful in a situation where students were opposed to answering questionnaires for some reason.

What might you do?

There are three possible ways of arranging for essays to be written on the course:

1. It may be required of some or all the students as coursework or project work, possibly as an alternative to writing an essay on the subject matter of the course.

2. It may be required of an individual. Most suitable for this purpose would be the class or faculty representative. If
one does not exist, a suitable student may be nominated.† His brief would be to informally survey the opinions of all the students before writing the essay, and he should be asked to provide a balanced view of student opinion.

3. It could be suggested to the students as a voluntary and perhaps optional activity, say, at the end of the course. This is, after all, the basis upon which questionnaires are used.

The disadvantages of the scheme can be lessened, the first by simply awarding an ungraded pass for submitting an essay. Alternatively, the essays may be submitted anonymously. The other problems, particularly that of the students not being able to gather their thoughts, may be overcome by presenting them (or the class representatives) with evaluation records of similar courses. A sample report appears in 2b, and the self-study courses issue checklist in Part III section 4 may also be useful. However, there then arises a trade-off between extending the range of the report by this means, and influencing it with outside perspectives, but this is something for you to decide.

Analysing the feedback

Again, I feel that it is reasonable that you should do more than simply read through the essays, and I would suggest that you make notes and factor analyse them as outlined in Part III section 6. If the number of respondents is low and the essays have not been anonymous, you may add to them by inviting the students to discuss their essay with you. Indeed, this may well form the basis of an interview (see Part II section 4).

Conclusions

What experience there is of this method is very promising. Students tend to make both criticisms and suggestions and the amount of feedback readily repays the limited effort required. An example of such an essay appears in 3b. It is slightly unrealistic in that it was prepared by a student with an eye to including the essay in a publication.

Other sections of Monitorkit providing useful information on this topic:

Part II sections 2, 4.
Part III, sections 1-6.

† The idea of using a class representative appears in several of the sections in Monitorkit, particularly sections 2, 3, and 4. If this is possible on your course, he or she may be involved in the monitoring strategy from an early stage.
Sample Essay

The following is an extract from a report written by an undergraduate student about her experience of the three Keller Plan courses she encountered in her undergraduate degree programme. It was written at the request of the teacher responsible for two of the courses, over the summer vacation after final examinations, and it was primarily intended for a publication.

"The first Keller Plan course I tackled was a Mathematics course designed to bridge the gap between 'O'-level and University entrance mathematics. I held a Mathematics 'A'-level already and thus the whole course was revision. However, I found the course quite rewarding in the end, although I found it hard to adapt to the Unit structure. Initially, I thought that this method of tackling the work was more difficult. Each point in the logic of differentiation was stated separately and for a large part of the course differentiation was from first principles only, which seemed unnecessarily complicated. However, I found that other students, tackling the material for the first time, welcomed this logical approach. Those students who more quickly completed Units, acted as tutors for the others and the course prospered on mutual aid. There were about twenty units to this course, although not all students completed them. When examinations came most students found that little revision was needed, because all the Units were built on the same foundations, and to pass to another unit the previous one had to be fully understood. After my initial scepticism, I was impressed with the depth and thoroughness of the course's teaching".

......"The next course (was) in my second year ..... I found the coursebook particularly difficult to read and this proved to be a great disadvantage. The concepts of the course were difficult to grasp, and this, with the text, made me disinclined to work as hard as I should on the course. This was encouraged by the pressures of other course work. Several of the other students, to whom Keller plans were new, found similar difficulties although some refused to work with an experimental course and wanted traditional lectures on the subject. I did not complete the course, only finishing about eight out of fifteen Units. However, I found I could easily answer the examination questions on the Units I had completed with very little revision. I could not answer questions on the other Units, although I had read through them several times before the examination. The problems and discussion of the Units, with the tutors at test taking, proved to play a larger part in successfully completing the Units than I originally anticipated, and the actual reading of the Units a lesser part."
4. DISCUSSION/INTERVIEWS - TEACHER AND STUDENT

Contents:

4a Background Paper
4b Interviews vs. Questionnaires
4c Interview Schedules

4a Background Paper

Introduction

The two most common ways of assessing student reactions to a self-study course are by interviews and questionnaires. Some advantages and disadvantages of both are set out in the accompanying list (4b), but it is probable that a combination of the two represents the best possible procedure. The discussion/interview is commonly seen as the most difficult measure for the teacher to use in evaluating his own course, yet informal discussions with students about the course happen frequently. The main values of the discussion/interview in this situation are:

(i) It enables students to state problems and issues as they see them.

(ii) It has the advantage of timing since, unlike questionnaires, it need not take several weeks to gather data on a particular subject.

(iii) It is responsive. You can respond to, and follow-up, newly emerging issues immediately.

However, interviewing by the teacher presents problems of role, status, and confidentiality. This may result in some doubt about the representativeness of the picture gained by such a process, and the procedures outlined below attempt to overcome these.

The range of possibilities

Discussion/interviews can be formal or informal; ranging from a well-prepared tape-recorded interview in the teacher's office on the one hand, to a chat in the coffee bar on the other. They range from highly structured to unstructured. Although both kinds must also be prepared for, an interview schedule is usually produced for a structured interview and the range of discussion may be limited. Conversely, the unstructured discussion/interview may aim to follow-up areas of interest generated by the student. It is simply necessary that they are prepared for and recorded (not necessarily tape-recorded!) in some way.

What balance is therefore desirable for teachers in self-study courses to overcome the above problems in discussion/interviews with students? Recent experience of this kind of activity indicates two possible strategies:
1. An on-going strategy

(i) Aim to talk to, say one student per week in an informal setting. It may be more reassuring to the student if this can take place in neutral territory, such as the corner of an empty classroom, or the coffee bar, rather than in your office.

(ii) Prepare for the discussion/interview by listing the issues you want to discuss. Part III section 4 will help in this. Ideas as to these issues will spring from your overall monitoring plan and from your experience of the course up to that point in time. This list should be turned into an interview schedule of a simple kind. Section 4c gives advice as to how this can be done. However, for this kind of informal interview it is probable that arranging the list into a logical order, and clarifying the way in which you intend to ask questions, will result in an adequate schedule. Take care in the way you phrase questions. Students will be particularly on the lookout for indications as to how you would like them to respond, so don't give the game away.

(iii) Invite the student for a 'chat about the course'. Ensure that, over a period of time, you talk to a representative sample of the class, both strong and weak, extrovert and introvert.

(iv) A suitable way of recording such a discussion/interview is by filling in the answers to questions, and the reactions to issues raised by students, on the schedule itself. This can be done during the interview, or afterwards. It is often of value to add questions, which spring to mind as you talk to one student, to the list before you interview the next. In this way the schedule should evolve to cover all the points which the students, as well as yourself, consider to be important.

To summarise, this on-going strategy involves a loosely structured informal discussion/interview pattern. It is probable that a written list of issues and reactions, added to throughout the course, would be all the preparation and recording needed.* This list would be useful for other purposes later, such as for planning the questions on a general course questionnaire.

2. Two formal strategies

In general, I doubt that formal interviews between a teacher and his students would produce useful, genuine information in return for a reasonable amount of effort. However, two types of people are exceptions. Firstly, the tutors on a self-study course might meet you for discussion/interviews regularly, either singly or as a group. They will have a lot of interesting points to make about the course, and particularly its materials.

Secondly, the possibility of using a class representative (existing or designated) has been tried successfully. He should be briefed to gather all the students' views on a number of issues prior to the interview. These issues may be left up to him, or they may be prespecified.

*(an adaptation of the recording method outlined in Part II 8a subsection 2 might work well.)
by the teacher, or he might be given a sample evaluation report (Part I, section 2b) or an issues checklist (Part III section 4) to focus upon. As the 'shop steward' he may have more confidence to face a structured interview, but his views must be treated with caution as the data is now doubly subjective.

**Tape recording**

The tape recorder is an excellent feedback device. (see also Part II section 8). If it is acceptable to your students, record your discussion/interviews sometimes. The playback can be very eye-opening. You see how much you misunderstood an answer, or angled a question in a particular way, and you can tell exactly what made the student clam up, or speak freely.

**Conclusions**

Using these methods, teachers can hold meaningful discussions/interviews with their students. It is a personal matter, and a matter of experiment what range of topics the discussions can cover. Some teachers find it easy to converse with students about the way they teach, others don't. Some teachers do fear that they will be prejudiced against students who are given the opportunity to criticise their teaching, others don't. The data which results from discussion/interviews is rich and immediate, and this process in particular has the effect of impressing upon your students your concern about teaching and learning.

Other sections in Monitorkit providing useful information for this topic:

- Part II section 2b, 6, 8, (particularly 8a - 2 for diary/notebook suggestion)
- Part III sections 1 - 6.
<table>
<thead>
<tr>
<th>Consideration</th>
<th>Interview</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Rate of response.</td>
<td>Good.</td>
<td>Can be poor.</td>
</tr>
<tr>
<td>4. Effect of topic (e.g. complex, personally sensitive, likely to arouse suspicion, etc.)</td>
<td>Sensitive.</td>
<td>Relatively insensitive.</td>
</tr>
<tr>
<td>5. Cost.</td>
<td>Staff time.</td>
<td>Staff time plus paper production.</td>
</tr>
<tr>
<td>6. Personnel involved.</td>
<td>Staff only.</td>
<td>Clerical assistance possible for early analysis.</td>
</tr>
<tr>
<td>7. Flexibility of instrument.</td>
<td>With slight modification, adaptable to individual situations.</td>
<td>Not adaptable.</td>
</tr>
<tr>
<td>8. Opportunities for catering for individual differences (e.g. speed adjustment, etc.).</td>
<td>Extensive.</td>
<td>Limited.</td>
</tr>
<tr>
<td>9. Opportunities for probing (following leads, etc.).</td>
<td>Possible.</td>
<td>Difficult.</td>
</tr>
</tbody>
</table>
What is an Interview Schedule?

The important part of an interview schedule is a set of questions you want to ask, or a list of areas you want to throw light upon. The schedule will probably occupy one or two sheets of paper and have space on it for your notes on the answers, and further questions and issues you wish to add to the list. It may also have notes reminding you of:

1. the reasons you wish to present to the students for the interview,
2. opening questions aimed at setting the student at ease,
3. the time and place of the interview. (These are optional).

How is it formed?

If you use the on-going strategy for discussion/interviews outlined in 4a it will be possible to build up an interview schedule as the course proceeds. At first the topics on the schedule will be dictated by the plan for course monitoring which you have devised (Part III sections 1, 2 and 4 will help in this). When you use the schedule with the first student you may find that his replies put new questions to mind, and show that others are irrelevant or impossible to answer. Gradually add these new questions to the schedule, and withdraw questions which are not producing useful information. In this way your schedule should evolve and grow to cover all the issues about the self-study course which both yourself and your students find important.

An alternative approach is to make the interview questions totally student-generated, by starting off the sequence of interviews with a general question such as: 'how have you found the course so far, what good features and what bad features spring to mind?' From the students responses it is possible to gradually build up a formal schedule.

A few tips...

1. Interviewing takes time. Don't try to cram too many questions or, worse, too complicated questions, into a short period of time. I find that 15 minutes is the minimum length of time for an interview to come alive. Hence, look at the questions and check that you can understand them and can see why they are being asked!

2. If students repeatedly misunderstand an apparently simple question it may well be because they do not think of the course in the same way as you. For example, this happened when I asked the question: 'how do your study habits develop on this course?' Eventually I found that these particular students never used the word 'study habit' nor were really conscious of having any such 'habits'. To get a useful reply, I had to rephrase the question: 'has your way of getting through the coursework changed during this course?' etc.

3. Your sole aim for the first few minutes of the discussion/interview should be to establish a rapport with the student, and put him at his ease. Reassure him about its confidentiality and its purpose, and ask straightforward questions.

4. During the discussion/interview, really pay attention to the student, and make it clear that you value what he has to say.
5. Don't always fill silences and pauses by talking yourself.

6. Many discussion/interviews don't warm up until towards the end! Capitalise on this, and return to early questions that fell flat or give the student a free hand to add any more comments he has to make.

7. Probably the best way to separate out fact from fiction presented for your benefit is by listening to tape-recordings of discussion/interviews. This is reasonably easy, particularly if you can get someone else to comment and criticise the questions and answers, as well as yourself.
5. DISCUSSION/INTERVIEW - TEACHER AND OUTSIDE ADVISER

Contents:

5a Background Paper
5b Example Outcome

5a Background Paper

Monitorkit faces one major problem in trying to help teachers to monitor (or judge, or evaluate) their own courses. The viewpoint of a staff member teaching a course in a new way tends to reflect commitment to the new method, and confidence in it; furthermore he tends to be preoccupied with day-to-day problems at the expense of more overall issues. This viewpoint is probably very important if the course is to really get off the ground, but it presents a problem when the teacher tries to stand back and judge the value of the course he is running. The other sections of Monitorkit try to get over this problem by suggesting procedures, providing materials, and raising questions which put the teacher in a better position to monitor the course, but this sections takes a different approach.

In this section I want to suggest that you, as a teacher of a self-study course, can be helped in monitoring your course by an outside adviser, who might, for example, have some experience of course monitoring, or developing a self-study course such as your own. However, most important, he would have an uncommitted, outsider's viewpoint of the course.

My suggestion involves the following:

1. You meet with the outside adviser only once or twice during the course. He is not always 'breathing down your neck'.

2. Your involvement with the adviser hinges around your monitoring of the course, and he may be provided with up-to-date information gathered by procedures outlined in Monitorkit.

3. The initiative is your own. If you think an outside adviser would help, you must contact a suitable person (see below).

4. This involvement may take the form of discussions, cross-checking, or interpreting the data you have gathered.

5. Alternatively, it may take the form of isolating certain questions which you feel that you, as teacher, cannot monitor. The outside adviser may then briefly continue his association with the course, and gather data to answer such questions (see example in 5b).

6. The essence of this scheme is that the outsider brings a viewpoint which complements that of the teacher, and helps in forming overall judgements on the course.
Procedures

Who can help you as an outside adviser? Two categories of people have been used:

1. Fellow staff members in, say, an education department, or even in your home department.

2. Research students, demonstrators, etc., working on the course, or supervised by yourself.

In particular, I myself have acted as an outside adviser on a number of courses and am always prepared to do so. If you would like to discuss this with me, please use the attached reply slip to arrange this.

However, I would like to encourage you to work with other categories of adviser. Help them by showing them Monitorkit, your monitoring strategy, and your results to date.

Conclusions

My experience of this procedure is quite encouraging. It does help the teacher get a better overall view of the course, and assists 'summative' monitoring as far as possible. It doesn't require too much effort on the part of the teacher.

Other sections of Monitorkit relevant to this method:

Part III sections 1 - 6

To: Will Bridge, I.E.T., University of Surrey, Guildford, Surrey, GU2 5XH

From: ........................................................

Phone: .........................

I would welcome discussions with you on the subject of my self-study course, preferably during one of the following weeks ..................

Please contact me and we can arrange something.

Signed: .............................. Date: ..........................
Example Outcome

This is an example of the kinds of outcome that co-operation between a teacher and an outside advisor can have.

In this instance I was invited to co-operate with a staff member using the Keller Plan. About 30 students were working on the course. The staff member had carried out a monitoring procedure, including issuing a general course questionnaire, but certain questions (marked Q) remained. The course was drawing to a close, and it was arranged that I should try to provide some answers (marked A) to these questions, by interviewing eleven of the students (of different abilities). Parts of the report which I wrote after the interviews appear below. From them it should be clear that an outsider can take a different stance in an interview, and ask different questions from a teacher.

1. Q. Two issues about the self-assessment questions (S.A.Q.s) had arisen from the questionnaire. Firstly there was a question about the extent to which the students used the S.A.Q.s as a guide for when to go for test. Secondly, the extent to which the S.A.Q.s were completed unseen was of interest, (both S.A.Q.s and solutions were in the units.)

A. Only one of the eleven students used the solving of S.A.Q.s as a measure of whether or not to go for a test. All but one of the remaining students did not do the questions completely unseen, and because of this they felt that the S.A.Q.s were not a good measure of when to take a test. The one student who did the S.A.Q.s unseen was the fastest of the group I talked to, and generally the weaker students indicated that they put less effort than the fast students into the S.A.Q.s before looking at the solutions. Two students treated the S.A.Q.s and solutions as examples, and rarely tried the questions themselves at all. However, the S.A.Q.s were used a lot; the students valued them, and none thought that the solutions should not be included in the units.

2. Q. The preknowledge list also emerged from the questionnaire as being of dubious value to the students, and this provoked a question about the extent and nature of use the list was put to.

A. All the students said they read the preknowledge list at the start of the unit. Half said they always knew all of the material required, and half admitted that it sometimes contained material they did not know. The former category of student made no further use of the list; however it was also true that four of the students who did not know the material said they went straight on with the unit in the hope that the procedure would explain it. For example, one student remarked... 'Oh yes, there were bits which I didn't know, but I didn't revise them, - I expected them to be explained in the Unit.'

The three students who took steps to learn preknowledge which they found they had not got did so by revising either
from the previous units (two students) or old lecture notes (one).

Clearly the preknowledge list was not being acted upon very much, with very few students covering any revision on the basis of it. A possible reason for this is that the list gives no guidance as to where missing preknowledge can be learnt (except in the references to previous units and these were followed up by students).

3. Q. Overall, the students responded that the course took them more time than the other (lecture) courses. Did time spent on this course come off other time allocations, or did this learning method encourage them to do more work?

A. Only one out of the eleven students felt that the course forced them to spend a long time studying, the remainder simply felt interested and encouraged to spend the extra time. For example, one student remarked "It took me about 3 hrs. per week, but that wasn't a disadvantage, it was just me, I felt encouraged to read the book."

The idea that the course encouraged broader reading came out a lot.

Two students found that they had to cut down on their time allowance for other courses as a result of studying this one. They were the two slowest students I talked to, and both said that was part of the reason why they stopped the course. The remainder found that the extra work was easy to fit in, although one student remarked "...but if all lecture courses took this much time it wouldn't be so good."

The most time consuming part of the units was universally thought to be the reading and understanding of the textbook, although those who did the S.A.Q.s more or less unseen found that they also took a long time. One of the students who had found the course taking away time from other studies said that this was in part due to the poor availability of the textbooks.

Overall the students who did feel that it was taking time away from other courses felt the course was worthwhile, but they simply were not prepared to spend the extra time on it since they spent little time on private study anyway. It is clearly very difficult to help this type of student.
6. DISCUSSION/INTERVIEW - TEACHER AND GROUPS OF STUDENTS

Contents:

6a Background Paper

Introduction

A straightforward and natural way of finding out how a class of students is reacting to a particular teaching method is to discuss it with them as a group. Generally speaking this can take place during classtime, and so it is important to ensure that all the students are present.

A group discussion has many of the advantages of the individual discussion/interview (see section 4): immediacy; student-centred nature; the possibility of following up, and generating new issues. It is also relatively easy to set up.

It probably should be recorded. The procedure suggested in section 4 whereby the teacher keeps an on-going list of questions, answers, and issues throughout the course would fit in well with this technique, as it is certain that both answers and new questions would arise during such discussions.

Advantages and Disadvantages

The class discussion has certain unique advantages:

1. The facility for gauging the overall 'average' point of view on any issue.

and 2. the facility for seeing the differences between groups of students (slow/fast, or young/mature, for example), and allowing students to discuss areas of disagreement.

3. It is possible that just one member of the group may recall an important, but generally forgotten, event; enabling more information to be got about it than would otherwise be possible.

However, there are some important disadvantages. Weak, or introvert students may find it particularly easy to remain silent in this situation and the conversations may be dominated by a minority. Even then, students may be understandably reluctant to really criticise the course in front of its teacher.

What might you do?

Experience of this method suggests that these problems can be overcome by introducing a degree of anonymity, and some activity into the discussions. In practice the following methods are possible:-
1. At the start of the group meeting, get each of the students to write out, say, a list of the advantages and disadvantages of the teaching method. Compile them into a class list on the blackboard (this can be quite anonymous). The discussions might revolve around the list, item by item, or area by area. A possible format for such a period would be to have the first half devoted to listing advantages and disadvantages, and discussing them; and the second half devoted to listing possible changes and improvements, and discussing them.

2. Buzz-groups may be used in the discussion. Split the class into groups of, say, four students, and give each group a task to discuss. The task might be to isolate the most difficult unit on the course so far, and say why it was difficult; or it might be to formulate a strategy to improve the course. After working in buzz-groups for 5-10 minutes, one member of each group should act as a reporter, and relate the group's conclusions to the whole class. The reporter usually feels able to be quite open about the conclusions, as he is not solely responsible for them.

Alternatively, the buzz-group may look at an issues checklist (see Part III section 4) or a sample evaluation report (see Part II 2b), and consider their relevance to the course they are working on.

3. Finally, a suggestion with a somewhat different emphasis. It is possible, and extremely useful, to involve a small group of students (or even just one student) in producing self-study materials. This may take the form of paid vacation work for an undergraduate student, or it might take the form of brief consultations with a previous year's students. They will have valuable contributions to make as to how the material may be explained or presented, and how the organisation may be made to run smoothly. Clearly they would have a far smaller contribution to make on the content or assessment of the course.

The first two kinds of monitoring can easily be worked into the routine of a course, for example they might be held regularly on the last class period of the first term. It can also be used at a crisis point, when you need to get to the bottom of a problem of some kind.

Experience of the use of this technique is encouraging, and although the results of discussions with a group of students should be treated with care and cross-checked where possible, a wide range of information can be gathered, and the students can be actively involved in developing a system of teaching.

Other sections of Monitorkit relevant to this method:

Part III sections 1 - 6
Part II section 2b, 4, and 8a subsection 2 (for diary/notebook suggestions.)
7. STUDENT PROGRESS DATA

Contents:

7a Background Paper
7b Weekly Progress Graph
7c Progress Profile Graph

7a Background Paper

Introduction

This section is concerned with what students learn on a self-study course, how quickly they learn it, and what helps or hinders their learning. Data on students' progress in learning the material of a self-study course is readily available, if unit test data is recorded. This built-in source of feedback, from Keller Plan courses for example, can be used in several ways.

There are three ways of collecting, and using, data about student progress and student learning. Of these, sections (1) and (2) are relevant primarily to courses, such as the Keller Plan, in which students take frequent progress tests prior to proceeding through the material.

1. Keller Plan unit test data can be presented graphically: two graphical forms are common (see 7b and 7c). 7b is usually used to keep an on-going record of each individual's progress, or the progress of the whole group. 7c is used to present the spread of achievement of the group at a given time.

Group data is important mainly for improving the course. Progress drops or pauses, shown particularly well on graph 7b, may indicate a need to improve, or shorten particular units. The effect of aspects of the course intended to help the students, or encourage them to work, can also be monitored in this way. Graph 7c, if produced, say every fortnight, can gauge the effect of a particular part of the course on the slow, the intermediate, or the fast students. It can also show where the students bunch up and work through the units at the same rate, or vice versa.

For presenting individual data, only 7b is suitable. However, if individual traces (see graph) are made public, there may be a danger that competition will take the place of co-operation amongst students, and the values of self-pacing may possibly be lost. If they are kept confidential, between the teacher and each individual student, it is probably wise to give the student some guidelines as to how quickly he or she should work. This latter approach does tend to ensure that the weaker students maintain a reasonable pace, although some fast students see the opportunity to relax their pace. Whether this record is kept solely by the teacher, or jointly with each individual student, it provides early feedback about which students need help, and which are potential failers, or drop-outs.
2. The completed unit tests provide useful feedback. In particular it is useful to analyse these to determine which units, and which individual unit questions, have been repeatedly failed. This information will be of particular use prior to revising the course material, to give an idea of which sections of which units require rewriting.

3. Overall examinations of student learning are a crucial way of monitoring the course. The final examination results are usually the acid test of a self-study course, for its teacher and the department he works in. Normally these are compared with results from a previous course taught traditionally, but sometimes they are compared with a parallel course taught traditionally over the same period of time (a 'control group' situation). Additional overall tests are also sometimes incorporated into an experiment with self-study courses, for example, pre-tests may be given to the students before they embark on the course, or retention tests may be held several months later, after the course has ended.

It is hard to give general advice about this kind of testing, much is best left to the subject expert - the teacher. However, some cautions and warnings can be given:

(i) Don't expect miracles. There has never been a university teaching innovation which produces statistically significant improvements in examination performance each time it is used, and in each place it is used. For many courses this is because the innovations don't teach (or don't only teach) what the traditional examinations examine. Be on the lookout for less overall, more unexpected outcomes, such as an improvement in the weaker students, or more students opting to take a similar course the following year.

(ii) Comparing the examination results of students on two parallel courses, one taught by a traditional method, and one by a non-traditional method, is full of problems. For example

Does the examination count (in terms of assessment) more in the traditional course?

Weren't the similarities (e.g. common textbooks, or common teaching staff) more important than the differences?

Were both courses really trying to achieve the same goals?

Were the students in both groups equally able? etc. etc.

(iii) What it is important to demonstrate is that examination results do not suffer as a result of the innovation. Most teachers end up with the finding that students do as well as, or slightly better than, students taught traditionally, and they look to other kinds of outcome to justify using an innovative method.

How much emphasis you place upon this kind of testing depends on two things. Firstly, how much importance you place upon learning as measured by a test of some sort, and secondly, the extent to which you are able to eliminate the problems mentioned above.

Conclusions

Much of the advice in this section 7 is straightforward and is
incorporated in many innovative courses anyway. Subsection 1 is so readily available on Keller Plan courses, and so valuable that I believe that it should be built into all such courses. Similarly, subsection 2 can be of great use to anyone revising self-study course units of any kind. Subsection 3, evaluation through examinations, is inevitable. Its advice may avoid disappointments, and produce findings of practical use.

Other sections of Monitorkit which may be of use for this technique:

Part III sections 1 - 6
Part II section 8a subsection 1
Part II section 8a subsection 2 (for diary/notebook suggestion)
8. RECORD KEEPING

Contents:

8a Background Paper
8b Recording Test-Tutorials

* All of section 8 is primarily of relevance to self-study courses only *

8a Background Paper

Introduction

This section is intended to help increase a teacher's awareness of what is going on in self-study class periods, and help him record his findings. Three kinds of recording are suggested. Firstly, that of attendance, secondly that of classroom activity, and thirdly, the tape-recording of test-tutorials.

Procedures

1. Attendance: Student attendance data on self-study courses can be most informative, since attendance at any particular class period tends to be optional, and students are able to 'vote with their feet'. Low attendance is probably the first sign of widespread procrastination. It may also point to a particularly difficult or uninspiring part of the course; or a heavy workload on other courses. A week-by-week 'head count' is worth keeping, particularly the first time a course is run, to pick out any problems as early as possible.

2. Classroom Activity: It may be important to be able to recall in some detail what went on in the self-study classroom from week to week. For example, this would be valuable when replanning the tutoring, the administration, and the accommodation of the course. It may also be useful to help make some overall, summative judgement about the course, and report this to your colleagues, or your course board! (see Part III section 1).

The kinds of activity which may be of interest include: when and to what extent queues form for tutoring; when tutors find themselves with little or nothing to do; times when weak students appear, or times when you suspect cheating is occurring, and so on.

I suggest such notes and records be written in a diary/notebook after each period. Indeed, this notebook might well also serve to record and develop the list of questions, answers, issues, and problems arising from other monitoring procedures, in particular: Feedback sheets (section 1), Discussion-Interviews (section 4), Group Discussions (section 6), Student Progress Data (section 7), and also 8a subsection 1. Such a notebook would present a very good way of structuring and recording a monitoring programme, and would be extremely useful in the following year's course, and when thinking about revising the course.
3. Recording Test-Tutorials: On Keller Plan courses (only) the 1:1 test-tutorial session is the primary contact between staff and students. Thus you may feel that it merits special attention. Here again the tape-recorder is a valuable feedback device.

With the student's permission (it is important that the student should be free to refuse), tape-recordings of test-tutorials between yourself and students, or other staff, postgraduate, or even undergraduate tutors and students, can be obtained.* The transcription of these recordings may be very valuable in improving individual tutors, the general way in which tutoring is undertaken, or the way in which undergraduate tutors are selected and trained. As you read through any such transcripts, (anonymous samples are available from the author), certain strengths and weaknesses become clear, and issues that can be worked upon, such as those listed in 8b, come to light.

Conclusions

Subsections 1 and 2 are obvious monitoring procedures, but they become particularly powerful when used in conjunction with a monitoring diary/notebook as suggested. Tape-recording Keller Plan test-tutorials may have important benefits to an established course, but it probably shouldn't be a high priority on a new course.

Other sections of Monitorkit which may be of use for this technique:

Part III sections 1 - 6
Part II sections 1, 4, 6, 7.

*A technical problem when doing this is background noise. I have found this less of a problem if the recorded sessions are held in the corner of the classroom.
Recording Test-Tutorials

The following issues have emerged from an analysis of transcripts of Keller Plan test-tutorials. Many of the problems can be identified, worked upon, and improved by the individual tutor himself.

Do you ask questions which tell students the expected answer?

Do you help them arrive at answers themselves?

Do you interrupt the students or inhibit them in any way?

Do you insist that the 'correct' answer is presented?

Do you pass on from correct questions without probing the knowledge behind it?

Do you provide information and/or help where it is needed?

Are you over-concerned with arithmetic/grammar etc.?

Do you invite the student to complete further calculations there and then?

Do you monopolise the discussion?

Is your tone helpful, bored, or aggressive?
PART 3  A STRATEGY FOR ASKING QUESTIONS

CONTENTS

A STRATEGY FOR MONITORING

1. Deciding on Aims
2. Devising a Programme
3. Using the Information
4. Questions to Ask

APPENDIX ON QUESTIONNAIRES

5. Writing Questionnaires and Feedback Slips
6. Analysing Questionnaires and Feedback Slips

CONCLUSION

7. Notes and Further Reading
1. DECIDING ON AIDS

First of all I must answer the question as to why I devised Monitorkit, and especially, why I devised it for innovatory self-study courses. There are three particular reasons why I believe it is important to monitor innovatory courses such as the Keller Plan, and self-study courses in general.

1. Such courses are new to both students and teachers. Neither can be sure of the roles which they should adopt, nor the procedures which will be necessary. This makes them feel concerned and unsure of themselves.

2. These courses have the potential to produce a lot of information about teaching and learning. The material is also tentative and flexible. In such circumstances it seems appropriate to gather and use the information so readily available.

3. Innovatory courses are hard work for the teacher, and sometimes for the students, so it is important to establish whether they are worthwhile, to find out what problems they have solved, and which new problems have arisen.

Course monitoring is a process of information gathering, followed by action on the basis of this information. Hence it is very important to list clearly at the beginning what kinds of information you aim to collect. Such a listing is useful throughout the programme. It suggests which measures to use for gathering the information, and which questions to ask. It indicates how the answers could be analysed, and how they can be used. The aims will certainly be modified and added to as the course proceeds, but they remain important.

It would not be possible to give direct advice about formulating aims, since they are very much a matter of personal priority. Part III section 4 suggests commonplace issues in self-study courses which you might aim to tackle. However, it may help to suggest one major category of possible monitoring aims.

**FORMATIVE** evaluation or monitoring aims to provide a feedback of detailed information about the materials and methods of the course, in order that it may be improved or 'debugged' as it goes along.

**SUMMATIVE** evaluation aims at a more overall view. Its purpose is not necessarily to improve the course, but to judge it, either by some absolute measure, such as its achievement of a list of learning objectives, or by means of a comparison with other teaching methods.

A mixture of formative and summative aims is common, but by and large I find that teachers using Monitorkit have more success in pursuing the formative aims.

To recap:

Formulate and write out a list of aims, refer to them periodically, and add to them throughout. They will indicate the questions to be asked, and the methods to be used.
2. DEVISING A PROGRAMME

Two Broad Strategies

There are no hard and fast rules for devising a course monitoring programme. However, two broad strategies are possible:

(i) Having previously decided what information the programme aims to gather, one strategy revolves around ensuring that each finding can be cross-checked. This is known as 'TRIANGULATION', and consists of a strategy where the same questions are asked in a number of different ways.

(ii) A complementary or an alternative strategy involves 'FOCUSED DEVELOPMENT'. In this kind of strategy the aims need not be so firmly pre-specified. Rather, as issues arise out of the day-to-day running of the course, they are gradually focussed upon, and monitored, until the mistake, disagreement, or misunderstanding at the bottom of an issue has been located.

A programme of course monitoring by teachers may require elements of either (or both) triangulation or focussed development, if the information gathered is to be valid and valuable.

Fitting in the elements of Monitorkit

The elements of Monitorkit spelt out in Part II must be fitted into the strategy and the aims which you decide upon. Below, I look at the eight types of data gathering procedures in Monitorkit one by one, and consider how they are best fitted into a course monitoring programme.

1. Feedback Slips. I have suggested that these tend not to ask wide-ranging, general or overall questions about the course. Rather they are applied to subsections of the course, such as individual coursework units or exercises. This means that they fit into an overall monitoring programme by providing detailed, specific information which is most useful for improving new or problematic parts of the course. This kind of information is often most valuable when the course is first run.

2. General Course Questionnaires. Since these questionnaires can ask both detailed, specific questions, and follow up more wide-ranging issues, they fit into an overall monitoring programme in a number of different ways. Used during the running of the course they can provide detailed feedback about the units, for example. Probably the most common use for them is as a summative measure, used towards the end of the course. Questionnaires produced for the latter purpose tend to be written on the basis of earlier elements in the monitoring programme.

3. Essays on the Course. Potentially, course essays may provide as wide a range of data as the General Course Questionnaire. However, unlike questionnaires, they are not certain to gather information about any particular part of the course. Hence, they might not, for example, be appropriate for a monitoring programme which was aiming to weed out the most time-consuming units, or disadvantages of the test periods. They would fit well into a programme which aimed to get the students' overall reactions to the course, and particularly their suggestions for improvements. This suggests to me that course essays may best be suited to the monitoring procedure on an established course, but this is by no
4. Discussion/Interviews. I believe that the teacher, unlike the neutral outside researcher, can't use discussion/interviews for in-depth summative monitoring; the inter-personal problems are too great. However, I believe they can play a role in a monitoring procedure similar to section 1, in keeping a finger on the pulse of say a new course, and gathering detailed feedback about it. They provide immediate, but not necessarily fully representative information, and may be more acceptable to students than a sequence of questionnaires.

5. Discussions with an Outside Adviser. A good place for this one-off measure is probably mid-way through the monitoring programme of a new course. This is because, at that stage, sufficient data and experience will have been gathered by the teacher for the discussions to be focussed and purposeful. Also, the teacher may be in need of support!

6. Discussions with Groups. Rather like section 3, this technique has the potential of producing a wide range of data about the course, but it may be hard to be sure of getting any particular item of detailed feedback. Thus again, it might fit appropriately into a programme which aimed to get the students' suggestions and priorities for improvements to the course. Alternatively, it might be held back from a monitoring programme, and only used if or when a problem which concerned the whole class (or a particular group of the class), emerged.

7. Student Progress Data. This seems to fit well into any kind of monitoring programme, since it provides valuable feedback both for and about the students. Because it is such an immediate record of the day-to-day progress of the course, it is probably of particular importance to a new course.

8. Record Keeping. The three procedures suggested in the background paper clearly fit into a monitoring procedure in different ways. The recording of test-tutorial is a direct response to the particular aim of improving test-tutorials. Attendance records and class activity records respond to similarly well-defined aims, not all of which are a high priority for new courses.

Any further recommendations as to how to devise a monitoring procedure must necessarily represent my personal viewpoint; however, perhaps they will help:

(a) In a programme of monitoring for a new course, I would consider the following:

(i) Student Progress Records - to keep watch on unit passing rates for individual students and for the whole class.

(ii) Either feedback slips or discussion/interviews - to perform formative monitoring on the course units in particular.

(iii) Possibly a general course questionnaire - I probably would not use both feedback slips and this general course
questionnaire, as I feel this would overload the students. The questionnaire would combine formative and summative questions, but would focus on how the course could be improved.

(iv) Finally, I might discuss the course with an outside adviser if I thought he could help the overall monitoring programme.

(b) In a programme of monitoring an established course,

(i) I would continue to keep student progress records because students change from year to year, and small changes to the units can have a big effect on the rate at which students progress.

(ii) Assuming that the course had settled down, I might use an essay technique to gather the students' view as to how it could be improved. Alternatively, I might want to focus upon, and develop, one particular area such as test-tutorials (through tape-recording), or the written materials (through involving students in revising the materials).

(iii) I might continue to issue one general course questionnaire, or hold one group discussion on the subject of the course, with the students each year. I would hope that these would keep me on my toes, and ensure that the course didn't stagnate.

Two additional comments. Just prior to revising all, or some of the course units, I would consider doing a test failure analysis (see Part II section 8), to establish which aspects of which units were causing the most trouble. Overall though, I doubt whether I would ever use more than two or three monitoring techniques on a course, and I might eventually settle for just one.

My final point concerns the mundane task of recording the data. For this I would tend to keep a single looseleaf diary/notebook of the kind mentioned in Part II section 8a. In this I would record everything about the monitoring procedure, starting from its aims and going on through its week-by-week findings and observations in such a way that, at any time, tentative or well-established trends could be seen emerging. In it, I would also include, for example, findings from questionnaires, and student progress data. This kind of record-keeping helps unify a monitoring strategy, and is invaluable when it comes to reporting the findings or revising the course.
3. USING THE INFORMATION

The ways in which teachers use the information which Monitorkit provides about their courses is so idiosyncratic that, again, it is impossible to do more than make certain specific suggestions and warnings. Four are particularly important:

**Two Suggestions**

1. The list of aims drawn up at the start of the monitoring programme should give a good idea about 'what to do with the data'. In particular it will pay to remember that the information can be used either in a formative way - to improve the course for next year, or next week; or in a summative way - simply to describe and judge its strengths and weaknesses.

2. When you do make changes on the basis of the monitoring procedure, check that they actually result in the intended improvement. Very many well-meant changes have unintended outcomes which are clearly not advantageous.

**Two Warnings**

1. Be wary of hearing what you want to hear. There is a great danger that Monitorkit could simply reinforce false prejudices. In the devising of aims, the strategy, and the detailed methods, I have included suggestions for reducing undue bias. However, the danger remains when it comes to interpreting, reporting or using the information. Perhaps cross-checking is appropriate here also. For example a colleague might be asked for his interpretation of some raw data, or for suggestions as to what he would do, given a certain finding.

2. There is a danger of over-reacting to unfavourable feedback (although the opposite danger is possible). This is particularly true where this is your first experience of direct student feedback. Students sometimes express all their dissatisfactions with all a department's teaching when they respond to a course-specific question.
4. QUESTIONS TO ASK

Introduction

In Sections 1 and 2 of Part III I did not spell out the aims that a monitoring procedure 'ought to have', as I believe that these have to be left to the individual teacher. Only then will he feel committed to take note of, and use, the data from Monitorkit. However, Monitorkit is primarily intended for self-study courses, and particularly Keller Plan courses. Experience shows that among such courses there is a common framework of questions and issues, some of which you might aim to answer in a monitoring procedure. They may point the way to questionnaire questions, or issues to be taken up at an interview.

The framework, covering most of the issues which arise in courses such as the Keller Plan, has seven headings. Many of the questions that are often asked are contained in this framework, but I would welcome your suggestions as to extensions or modifications of the framework. The headings are:

1. The Unit Structure and Content.
2. Tutorial Periods and the Testing.
3. Workload.
4. Assessment.
5. Motivations, Attitudes, and Difficulties.
7. Summary and Follow-up.

The Framework

1. The Unit Structure and Content

1.1 How (a) enjoyable, (b) easy to learn from, (c) useful to this course, and (d) useful to other courses, were the following components: preknowledge list; objectives; procedure; textbook; essay or report paper; lectures or films; the Self-Test Questions.

1.2 Unit Structure: were the units too prescriptive, or, conversely, were the students not clear what they had to do?

Unit Content: was the content relevant to their studies, interesting, and at the right level?

Unit Length: were the units too long, too short? Did the students get a fragmented view of the material?

(The above questions (1.2) may be asked for individual units or for all the units as a generalisation.)

1.3 How much use was made of supplementary work such as laboratory and computing exercises, and what value did the students place on these?

1.4 Books: what was the amount of use, and perceived value, of (a) the course textbook(s), (b) other recommended texts, and
(c) books on the subject, not specifically recommended? Also journal articles.

2. **Tutorial Periods and Testing**

2.1 **Tutors**: what do they actually do? how and at what level do they mark tests? what is the student reaction to being tutored and to the individual tutors? If undergraduates are used as tutors, how does this affect their progress? What degree of staff/student contact exists? Is this valued?

2.2 **Tests**: student reaction to tests, particularly

(i) apprehension
(ii) value or otherwise of feedback about learning.
(iii) perceived level of mastery.
(iv) relationship between the test and the unit.

what actually happens during a test-marking session? what is the quality and quantity of feedback from students gathered during a test session about the course and the units.

2.3 **Test Room Conditions** - noise, crowding, variation of attendance, availability of staff or tutors for general tutoring on the course.

3. **Workload**

3.1 The amount, and variation of, time spent by students on the course. Students' attitudes to the workload.

3.2 When and where the work is done.

3.3 Does work on this course encourage or develop particular ways of studying? If so, what are they?

3.4 Do these ways of studying transfer to other courses? Does the workload affect other courses?

4. **Assessment**

4.1 **Unit Progress Data** - what does it tell us, about particular units and particular students? Is it related to examination performance? If some units are optional, which are the most popular and why?

4.2 Can, and do, students cheat the system?

4.3 How do the students fare on the final examination, later examinations, retention tests, etc.? Are any groups or types of student particularly helped by self-study teaching methods? How does learning by this method compare with learning by other teaching methods?

5. **Motivations, Attitudes and Difficulties**

5.1 **Motivations**: Do the students feel there is adequate interest and pressure to work? Would they like more?
What parts of the course make them keen to work? What parts of the course put them off? How is their interest in the subject matter affected by the course? Is the course seen as a competitive or a co-operative venture, by slow and fast students? Do optional activities, such as 'stimulus' lectures have any effect on motivation?

5.2 Attitudes: Development of attitudes towards: working on self-study material; self-pacing; really mastering material before progressing; tests and continuous feedback/assessment.
Do the students feel that the course is an efficient/enjoyable means of learning? To what extent do the students feel involved in the running of the course, and do they wish to be? What attitudes towards the method and content of the course do the staff reveal?

5.3 Difficulties: Do students (particularly foreign students) have language or reading difficulties? Does the course produce logistical difficulties, or difficulty in timetabling and staffing?

6. Institutional Factors

6.1 How does the course fit in with (a) later courses dependent on it? (b) the rest of the degree structure? (c) the examination structure? (d) the tutorial structure?

6.2 Do the following people affect the course? (a) The head of department, (b) course boards, (c) other colleagues?

7. Summary and Follow-up

What are the overall likes and dislikes of the participants? What are the most important effects of the course on the students?

Would the students like further courses organised in the same way? What types of courses are most suited?

What effect has the evaluation had on the course?

After the course, is the material used again in later years?

How do the students who did not cover all the units get on in later courses? (particularly if they build upon the self-study courses).

What are the overall logistical demands of the course, in terms of staff cost, duplicating, administration, etc.?
5. WRITING QUESTIONNAIRES AND FEEDBACK SLIPS

Introduction

It is likely that the sample questionnaires and feedback slips in Monitorkit will not fit in exactly with your monitoring strategy, and therefore you may want to produce your own. Much has been written on this (see Part III section 7), but I find the following points of value:

1. Formulate Aims: Before you write out a questionnaire or feedback slip, you should be clear of what you are hoping to achieve, and what information you need. Write a list of your aims, and the categories of information you wish to gather.

2. Check Aims: Check through this list to ensure that the aims are straightforward enough to be asked in a simple questionnaire.

3. Formulate Questions: Translate the list of aims into a number of questions. Take care not to (a) angle the questions without realising it, (b) restrict the range of possible answers to your preconceived ideas, and (c) make all the responses of the same kind.

A variety of formats can be used for asking questionnaire questions:

(a) Open, non-directive - e.g. 'how do you find this method of learning about alternating currents?'

(b) Open, directive - e.g. 'Did you find any part of the units particularly useful?'

(c) Rating Scales
   e.g. (i)'Please rate the following units on a five point scale:

<table>
<thead>
<tr>
<th>Unit 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>

   (ii)'Please ring the number most nearly corresponding to your reaction to the following statements.'

   'The course is very hard work'
   Strongly agree       Strongly disagree
   1 2 3 4 5

(d) Comment - e.g. 'Please comment on your rating of each of the units.'

(e) Justify Questions - e.g. 'If you responded ... above, please justify this.'
(f) **Response Selection**

  e.g. (i) 'Please ring the amount of time you spent on Unit 1: '
  
  0 - 2 hrs.  2 - 4 hrs.  over 4 hrs.

  (ii) 'Would you choose to take another course taught in a similar way ... yes/no (delete one).'

(g) **Checklists**

  e.g. 'Which of the following books have you used on this course?'

  Karl Marx, Capital  [ ]

  I. Fleming, James Bond  [ ]

  The Bible  [ ]

  (tick as appropriate)

**Selection of Format** (see also Part II section 2)

Some suggestions can be made about how to choose between these formats. Rating scales are mainly useful for comparison purposes, for example to detect if one of a series of several units was found to be particularly time-consuming. To help this comparison, rating scales are often grouped together. However, it is probably wise to vary the format of the questions occasionally in order that students do not get into a routine of answering questions without thinking. Finally, the tendency to group closed questions at the start of the questionnaire, and open questions at the end, should be avoided, firstly because the students tend to tire of filling in even a short questionnaire, and tend not to make comments at the end, and secondly because at the end their comments may be limited to, and directed by, the areas covered by the questionnaire.

4. **Dummy Run**: After drafting the questionnaire or feedback slip, you should cross-check that it will produce the data you want. One way to do this is to give two or three students the draft to fill in. Sit with them during this dummy run, and get them to interpret each question, and tell you if any questions are hard to answer. This should give you a picture of how the class will interpret the questionnaire and respond to it. If this is satisfactory, the questionnaire or feedback slip can be distributed.

5. **Getting them Back**: Once the questionnaires have been distributed to the whole class, a common problem is that of getting good return rates. If you are prepared to send out reminder notes, and generally chase-up students, you might get 60-90% return rates. However, there are at least two other approaches. Firstly, if for example, feedback sheets are a regular feature of the course, then students might only be allowed to take the test for a unit when they had handed in its feedback slip. Similarly for laboratory or computing feedback slips. Getting marks, finishing one exercise, or going on to the next may be made conditional on handing in a feedback slip. A second alternative is to take some opportunity when the class is all in one place (for example, at the end of a lecture, or a laboratory period), to hand out and collect general course questionnaires on the same occasion.
As in Section 5, this is a topic that has often been written about, and some commonsense rules emerge.

1. Aims : Go back to the list of aims you drew up at first. These will provide a starting point for the categorisation of data in the questionnaires.

2. Rating Scales : Closed questions, such as rating scales, response selections, and checklists, should be processed numerically or quantitatively. Except for rating scales, this involves first ascribing a numerical value to the responses, for example:

   'Please ring the amount of time you spent on unit 1.'

   0 - 2 hrs.  2 - 4 hrs.  over 4 hrs.
   (1) (2) (3) ......ascribed numerical value

   Useful information from this kind of question can be obtained by calculating the class mean value, and relating it back to the question. As stated previously, it is in the potential for comparison amongst rating scales which gives them particular value. Hence, for example, the class mean value for the above question might be compared with that obtained for each successive coursework unit to find out which units students find particularly time-consuming.

   An extension of this analysis is to plot a histogram of the responses to each rating scale, since mean values often hide bimodal distributions, for example.

   In each case, the analysis is trying to determine trends such as particularly weak units, or growing trends of opinion.

3. Written Comments : Written comments tend to take longer to analyse. They can be processed in many ways, but it is important to do more than just read through them, for this encourages selectivity and bias.

   All the responses to each written comment question should be gathered together on a single sheet of paper (this might be done by an assistant). Some qualitative factor analysis can then be applied. This can be done by reading through all of the responses, and working out the major categories or headings under which responses have been made. For example: 'Gave up after 5 hrs. ...', 'Took ages to complete this unit', and 'Was put off by the amount of reading it recommends', might all be categorised under the heading of 'Heavy Workload'

   There are theoretical problems about this very simple approach, for example rating scales are often not composed of equal steps. Section 7 suggests sources where this may be pursued if it is of interest. Overall, I believe they pose a problem for educational researchers rather than teachers involved in course monitoring.
Eventually you will be left with a number of reply categories for each question, and this should include an indication of how many students' responses fell into each category. For example:

**Question:** "Overall, how do you think the course could be improved for next year?"

**Answer:**
(a) More lectures (75%)
(b) Less Mathematics in early units (42%)
(c) Less Mathematics altogether (28%)

Other responses regarded as particularly important:
"Stop repeating material covered in Dr. Black's course."
"Take more notice of us Civil Engineers."

4. **Cross-Check:** When you reach the stage of drawing out trends from numerical questions, or categorising the responses from written questions, it may be important to cross-check your interpretations of the data. You can do this by asking a second person, not involved in the course, to check these interpretations, and decide whether the categorisation is representative of the comments made by the students.

When you have gone through these stages, write out the results as you see them, noting both judgements which you have come to, and information which you feel is missing. This advice, and the suggestion that you should analyse the questionnaires as soon as possible after you collect them, is presented because I know from experience that it is very easy to put off the analysis. Furthermore, even when it is done immediately, unless you note the judgements which seem to emerge clearly and obviously from the data, they will soon be forgotten or distorted.
7. Notes and Further Reading

The second edition of Monitorkit was prepared in September 1975 on the basis of field-trials of an earlier package during Spring and Summer 1975. It is part of a programme aimed at involving teachers in research and evaluation on courses they teach. More information about the programme is available in two unpublished papers, available from the author:–

"The Past, Present and Future of Monitorkit".
and
"Microevaluation - a new emphasis for small scale programme evaluation" (jointly with D. Laurillard).

The field-trials are continuing, and I would welcome the comments of any readers on the subject of Monitorkit, or the rationale of the programme.

Further Reading

This package covers many well documented areas in exceedingly brief, lay terminology.

Three publications are suggested for those who wish to go further than the limited scope of Monitorkit:


Parlett, M., and Hamilton, D., "Evaluation as Illumination : a new approach to the study of innovatory programmes", Occasional Paper No. 9, October 1972, University of Edinburgh. (This paper characterises a movement away from examining and testing by experts as the sole means of educational evaluation; since Monitorkit is rooted in this movement, the paper may be of interest.)