A COMPARISON OF PROGRAMMED LEARNING WITH CONVENTIONAL
LECTURING IN THE TEACHING OF SPECTROSCOPY AND IONIC
EQUILIBRIA TO UNDERGRADUATES.

A thesis submitted by

ALAN W. MACKAILL, B.Sc., Ph.D., F.R.I.C.

in part fulfilment of the conditions for admittance to the
degree of

MASTER OF EDUCATION

UNIVERSITY OF EDINBURGH                                    1974

# CONTENTS

# SUMMARY

An investigation, comparing the effectiveness of formal lecturing with programmed learning, in the teaching of first year undergraduates reading chemistry is described.

In Chapter 1, problems facing early workers, such as Skinner and Crowder, are discussed and the relevance of programmed learning to tertiary level teaching considered. Recent advances in the teaching of science subjects are described, particular attention being paid to the use of programmed methods in chemistry. A review of comparative studies in other disciplines is made with critical comment on experimental designs used. Finally suggestions are made concerning the use of experimental criteria in comparative studies.

The second chapter describes the experiments which were carried out but, prior to this, discusses the aims and scope of the research, dealing with such aspects as the selection of the sample, internal and external validity, and the hypothesis to be tested. The problem of experimental design is dealt with at length and details of the two designs chosen (namely Latin square and randomised group) are given. Experiment 1 is then described. Details concerning the topics, teaching methods and tests are given and the results dealt with using analysis of variance. Experiment 2, while of different design from Experiment 1 was carried out to

test the same hypothesis the statistical technique in this
case being analysis of covariance.

Finally analysis of the post-test is carried out
by assessing (i) how marks are distributed for experimental
and control groups (chi-square) and (ii) which particular
behavioural objectives are measured by each question.
Suggestions are put forward to explain the experimental
results in terms of post-test sensitivity.

The concluding chapter consists of a detailed
discussion of research into teaching methods concentrating
particularly on how results should be interpreted and used.
It is suggested that considerable further study is required,
and those areas where such study is likely to be fruitful
are identified. Finally, it is pointed out that there
remains a number of unsolved problems and suggestions for
the future solution of these are put forward.

# CHAPTER 1

'If, by a miracle of ingenuity, a
book could be arranged that only
to him who had done what was
directed on page 1, would page 2
become visible and so on, much
that now requires personal
instruction could be managed by
print.' - Thorndyke and Gates.

In actual fact no such miracle was necessary.
What was really required was revision of traditional text
book format, a method which would ensure the active
participation of the reader.   Such a method has now been
worked out which, for want of a better title, goes by the
name of programmed learning.

When Thorndyke was dreaming of 'a miracle of
ingenuity' the prototype of a teaching machine was being
experimented with, and tested in an American College.
As is so often the case with many inventions the
significance of this one was not realised at first.  It
was simply regarded as an additional device to be added
to the growing number of such devices, for which American
patents had been applied.   There was indeed, nothing to
suggest that this particular technique was in any way
different from any of the others.

The concept of automatic teaching evolved slowly
from the 1920's to the mid-fifties when its development
suddenly accelerated.   The effects were felt throughout
the whole educational world.   Today programmed learning
is a branch of study in its own right, with its own
extensive literature, techniques and technology, its
famous personalities and its rival factions.

Unfortunately the prospects opened up by these developments have for some teachers seemed startling and extravagant.   To many, the literature and language of the subject seem unfamiliar and incomprehensible.   It would, however, be a mistaken belief for a teacher to assume that the principles of programmed learning are too difficult for him to master;   there is no reason to suppose that these principles are the province of the learning theorist, the electronic engineer or the cybernetician.

## Pressey's early work

Credit for the invention of programmed learning is usually given to Sydney Pressey, a lecturer in Ohio State University.   Strictly speaking, however, the invention belongs to no particular individual.   Initially Pressey was concerned with a way of saving time in the scoring of objective tests which he had to administer each week to large numbers of students.   Pressey estimated that he spent one thousand hours per year marking such tests and eventually designed a machine in which the student pressed a key.   The response was recorded by the machine as correct or wrong and the student's total score was available at the end of the test.   The machine was subsequently modified in such a way that items could not be presented to a student until the previous items were correct.   Much of Pressey's work, while not unnoticed, did not impress his colleagues and little progress was achieved in the development of programmed learning. Writing in 1932 Pressey said,

'There has been, so far, relatively
little development of instruments
specifically for the very extensive
yet analytical research typical both
of modern educational investigation and
also more general of the social sciences.
New instruments greatly facilitating
research may soon appear.    There may
then be sweeping research advances in
these fields.' (1)

These advances were not to materialise for many
years and when they did come they were the spin-off from
research in military requirements, occasioned by World
War II.    Both B.F. Skinner and Norman Crowder became
interested in programmed learning while working on
military projects.   Working independently and on different
lines both Skinner and Crowder laid down the foundations
of what we know today as programmed learning.    Both men
considered conventional classroom teaching to be inefficient
and both were dissatisfied with average text-book
presentation.

## Skinner's work - the linear programme (2)

In his early laboratory work with pigeons, Skinner
achieved remarkable success in training these birds to
exercise various dancing feats such as waltzing round
in figures of eight.   Basically Skinner made use of the
same methods which all animal trainers use:  namely,
conditioning.   Each pigeon was rewarded with food when
the correct movement was executed.   In an effort to
apply such methods to the learning process, with the same
degree of control that existed in the laboratory

situation, Skinner devised the technique known as linear
programming.

The linear programme consists of a series of
connected items or frames each one related to that
preceding it and related also to that following it.
Each frame is incomplete in the sense that it requires
the learner to do something (usually answer a question)
before the next frame can be dealt with.  The questions
are usually easy and so as the learner works through the
programme, he receives a series of stimuli which reinforce
the learning as feelings of doubt and insecurity are reduced
and confidence increased.

The well known cliche 'nothing succeeds like success'
illustrates this situation in the sense that the stimulus
of knowing that he has answered the question correctly,
pleases the student and encourages him to try the next
frame.

## Crowder's work - the branching programme (3)

To those who do not accept the behaviourist
philosphy, the fundamental fault in linear programmes
is the fact that they do nothing for the student who does
make a mistake.  They would assert that the good teacher
does not dismiss the wrong or partially correct answer,
he rephrases the question, or tries a totally different
one in order to make the learner improve on his initial
performance.  This pupil-teacher interaction was built
into programmed learning by Crowder by instructing the
learner to turn to specific pages for each of the possible

responses to the question. Those who are correct press on as in linear programming; those who are wrong are provided with more information and questions to clarify the situation or perhaps they simply find a bald statement telling them they are wrong and to answer the question again after more careful thought.

It was because of this considerable side-tracking that the description 'branching' has been applied to Crowder's system.

To date, both Skinnerian and Crowderian programmes have been widely accepted in educational circles, and though there are many differences between the two techniques they are essentially complimentary.

While it is true to say that programmed learning, using either of the two techniques described above, has received wide acceptance, it is important to note for the purposes of this present study, that the acceptance has been mainly at the secondary school level. Programmed learning at the tertiary level has undergone relatively slight development; confined to a few specific subjects such as medicine, chemistry, some foreign languages and a few branches of physics (4).

One reason for the lack of interest in this method of teaching may be that, at the tertiary level, particularly in the sciences, knowledge is expanding almost exponentially and consequently the contents of a programme are soon out of date. Since it is estimated that a programme which would take a student three hours to work through, takes something of the order of eighty hours to prepare

a relatively long life for a programme must therefore be envisaged if the compiler's time and energy are not to be wasted.

This is the principal reason for the relative lack of enthusiasm for programmed learning in tertiary centres, particularly the Universities.   Universities have consequently devoted their teaching energies to the traditional modes of instruction, namely the lecture, tutorial and laboratory, since the content material may easily be up-dated.

This is not to say that these methods do not have their faults.   Lecturing as a means of teaching is now so often attacked (particularly by students) that some justification is really necessary for its retention.   The principal arguments against its use are:

    (i) that it promotes passive learning which is
        less effective than activity learning.

   (ii) students have little opportunity for asking
        questions as soon as a difficult point arises.

 (iii) students are all forced to receive the same
        content at the same rate.

  (iv) students are exposed to one, usually biased,
        interpretation of the material.

   (v) lectures are very often dull and badly
        presented.

A number of researches carried out within the last ten years (5) (6) shows students to be less condemnatory of lecturing than was hitherto thought to be the case. Generally speaking science students consider the lecture to be a useful method for introductory material or for

material not yet in text-books.    Science students do not
look for originality which seems to be important for
arts students (6).    However, medical and dental students
feel that they have too many lectures.

As might be expected lecturers tend to favour the
lecture as a teaching method.    Lecturers in scientific
disciplines feel that the lecture is the best method to
present new material which the student cannot master
unaided, and as already pointed out, lecturers' notes are
easily up-dated.    The criticisms by students (and some
educational psychologists) notwithstanding, most lecturers
claim to cover the syllabus with the minimum amount of
illustration and claim that they are in a position to modify
their approach, or their material, should students require
it;  a possibility denied in a teaching machine, however
sophisticated.

The principal advantage of the lecturing method is,
of course, the fact that a large number of students can be
taught at one time whereas the laboratory or tutorial class
is very much smaller (7).

Recent advances in tertiary science education

The lecture remains at present, the predominant
teaching method used in tertiary science education, and
yet very little research would seem to have been conducted
into its function, either by scientist or educationist.

It is generally accepted that, for a variety of
different reasons, not necessarily connected, lecturing

is disliked by some and advocated by others, within both
staff and student bodies.    Changes are apparent, however,
in certain quarters, which are due, in large measure, to
the introduction of 'self instruction ' systems.

For some time now many lecturers have issued printed
or duplicated notes which could easily be modified and
up-dated.    The belief behind this system was that information
possessed by the student at the end of the lecture was complete,
free from error, (certainly not the case when the student
makes his own notes) and thus produced more efficient learning.
The argument against this is, of course, that the student
actually learns by taking his own notes.    It does seem,
according to McLeish (8) that providing students with printed
notes does improve learning.

Elton et al (9) have carried the provision of notes
a stage further, by making tape recording of lectures and placing
these in the library within two or three hours of the lecture,
along with duplicated material which extended the printed
lecture notes.    The obvious criticism here is that attendance
at lectures might suffer.    According to Elton, this did not
prove to be the case, and even if it did, it would not matter
much.    Elton decided to replace a number of lectures in the
course with taped material and visual material (the so-called
tape/slide presentation) and found that the students did
equally well in the examination no matter which teaching
system was used.    An important discovery was made, however,
namely, that the students and staff did not like the tape/
slide method.    The principal reason for students attending

lectures is that this is the only way to discover what the boundaries of the course are. The student may not learn anything by attending the lectures, but he discovers what he is supposed to learn. When the examiner and the lecturer are one and the same person attendance at lectures is the best way to find out what is likely to be set in examinations.

The general view for both staff and students is that lectures are popular and efficient teaching instruments. Progress in tertiary science education, according to Elton, will only be made if the lecture method is not replaced but rather supported by new techniques.

More recently, however, Elton has modified this point of view (10). A quite new approach known as the Keller plan, which abandons the lecture totally has been developed in the U.S.A. (11). The students are simply given units of work in written form which states the objectives to be achieved by the end of the unit. Various means of achieving the objectives are given, such as reading references in text-books, additional notes, and problems for solution. When the student thinks he has mastered the unit he is given a test of twenty minute duration. As soon as the test is completed it is marked and the results discussed, between student and tutor. If the student passes the unit test he proceeds to the next unit, but if he fails he must do further work on the unit and take the test again. Considerable use of this system is now being made in M.I.T. and other centres of tertiary

education (12). When this system was introduced in the
University of Surrey (10) the vast majority of students
were enthusiastic. Elton says:

> 'Contrary to what might be expected, students
> did not feel that this method tended to spoonfeed
> them; rather the opposite, they found that they
> were given much more freedom to work within the
> broad framework provided by the units. They
> commented that their time was being used more
> efficiently and that they worked harder than
> they would have done under the traditional
> lectures and tutorial method; that they learned
> more; that they understood the course while
> they were going through it; that they had closer
> contact with the staff than in other courses;
> and that they would welcome other courses given
> on the Keller plan method.'

## Programmed learning in chemistry

Although the student enthusiasm for the Keller
plan reported by Elton has never as yet been in evidence
for programmed learning, there are nevertheless many
similarities.

Within the subject of chemistry the earliest
contribution made to programmed learning was by Gunstone
and Moyes (13). Further developments were made by
Hoare and Inglis (14). In 1966, a conference was held
at Oxford under the auspices of the Royal Institute of
Chemistry and the Chemical Society on the use of programmed
learning in the teaching of chemistry to degree standard.
At this conference three topics, principally, occupied
the attention of the speakers (both educationists and
chemists). These were:

> (i) the usefulness of programmed learning as
> a teaching method in courses to degree standard.

(ii) the difficulties inherent in the writing of good programmes.

(iii) the most efficient use of the resulting programmes in view of the time and effort spent on them.

Originally much doubt was expressed by chemistry lecturers about the usefulness and appropriateness of programmed learning in chemistry at degree level but criticism is now much less vociferous than at first. It is now agreed that in degree courses, students should be encouraged to abstract useful information from a variety of sources and that in this context, programmed learning is useful in that it teaches students the value of having material logically organised. Most lecturers take the view that programmed learning has great value in service teaching, since it is felt that where motivation to learn chemistry in, say, a medical course is small, programmed learning can increase understanding. This kind of course is the only one in which much research has been carried out (14). There is a definite reluctance to make use of programmed learning in honours courses due perhaps to lack of good programmes and perhaps due, as already suggested, to the fact that programmes are difficult, if not impossible to up-date;  a very important aspect of teaching at honours level.

An important difficulty in writing programmes for undergraduates, is that from the economic point of view, it is desirable that a programme devised in one university should be useable in another.  Unfortunately undergraduate

courses are such that emphasis on a given topic varies
from one university to another.

The writing process does not end when the topic
or subject area has been covered.  The programme has
to be "tried out" or pre-tested and its faults identified.

Clearly, then, the work involved in constructing
good programmes is considerable.  The Oxford Conference
suggested that in view of this, a central organising
body should be set up to:

    (i) circularise programmes for testing

    (ii) delegate universities to particular topics
           to avoid duplication of effort.

  (iii) organise meetings to discuss results and
          train lecturers in the writing of programmes.

As might be expected, rather more interest has been
shown in programmed learning in the U.S.A. than in this
country.  The American equivalent of the Oxford Conference
took place three years earlier in 1966 under the Chairman-
ship of J.A. Young (15).  Generally speaking, very
similar points of view were expressed in the reports from
this conference, to those in the Oxford Conference.

The principal mistake made by these early workers
in programmed learning seems to have been an almost
universal belief that it was only a teaching aid.  Very
little attention seems to have been paid to the possibility
that it might also be a learning aid.  In other words,
little research has been done to discover the student
point of view.

The most important contribution in this field has

been that of Hogg of the Department of Chemistry,
University of Aberdeen.   In his first investigation
of student opinion of programmed learning (16) Hogg
administered a questionnaire of forty-seven items to
his second year students.   The questionnaire contained
both closed and open-ended items.   The responses were
analysed with the students classified into four ability
groups (of equal size) based on probable degree class,
but no marked difference of opinion was detected between
the groups.   Hogg does not say whether any attempt was
made to consider within-group differences which suggests
that analysis of variance was not included in his
calculations.   A favourable response to programmed
learning was obtained from 80% of the students.

It should be pointed out here that Hogg was not
using programmed learning as a teaching device but as a
learning device;   in other words, none of the usual
teaching methods had been abandoned, the programmes being
issued to students to study at home as an adjunct to their
normal course.   Had the same students been asked to report
on the same programmes used as an alternative to lecturing
the response might have been quite different.

More recently Hogg (17) has extended his researches
to other universities and tertiary education centres in
the U.K. giving a much larger sample of students.   (516
respondents from 16 institutions).   There had been a
little evidence from Hogg's first project that academically
weaker students rated programmed learning higher than did

the better students.    This, however, was not found in
the larger survey.    Nor was there any difference of opinion
between chemistry students and students reading other
science subjects.    The proportion of students in this
second survey, responding favourably to programmed learning
was rather less than in the first (69%).    Again, however,
the technique was used as an adjunct to conventional
teaching and not to replace it.    Generally the students
found the programmes to be of moderate interest and
usefulness.    There was a marked preference for programmed
learning to be used for private study particularly where
it covered ground already dealt with by more conventional
methods.

The reaons for this are not given but it is not
impossible that, after a one hour lecture or tutorial
(perhaps badly prepared by the lecturer or badly delivered),
the student is left with only a vague idea of the subject
matter discussed, or incomplete lecture notes, and
consequently feels a strong need for additional material,
delivered at a different rate in a different way.

Comparative studies

It would appear that, with the exception of Elton (10)
the majority of science lecturers consider programmed
learning to be of use only as an adjunct to the more
conventional methods.    It was certainly not as an adjunct
to conventional methods, that such innovators as Skinner
and Crowder, devised programmed learning.    As far as they
were concerned programmed learning was decidedly a

replacement for, and not an adjunct to, conventional teaching.

Given that conventional teaching is inefficient how can educationists be sure that the proposed alternative is any better ? The problem can be solved by conducting a well designed experiment to compare method A with method B.

For many years now, multifarious experiments have been carried out to determine which of several methods is the best way of teaching French, or Algebra, or whatever. Unfortunately, when one or other method became popular it was not always due to the collection of good experimental evidence. As Cronbach has pointed out -

> 'In education, unfortunately, there is great
> furore about whatever is announced as the
> latest trend, and the schools seem to career
> erratically after each Pied Piper in turn.
> This giddy chase keeps them almost beyond
> earshot of the researcher standing on his
> tiny, laboriously tamped patch of solid ground,
> crying in a pathetic voice "Wait for me; wait
> for me." (18).

Results obtained from comparison studies have been conflicting and not readily generalisable. In some situations conclusions arrived at can only be applied to particular students in certain particular situations. In other situations where two methods were compared the experiments failed to take account of differences between teachers. For example, Mr X might be good using Method A and poor using method B, whereas his colleague Dr Y is just the opposite. Only a design

which allows for change-over of teachers and classes can take account of such differences.

Comparative research in the U.K. has tended to be of two types:

(1) Comparison of programmed learning, using teaching machines with programmed texts,

(2) Comparison between programmed techniques and formal teaching or lecturing.

Three papers published simultaneously in 1963 indicated clearly that formal teaching and programmed methods are equally effective. In the first of these three studies Wallis and Wicks (19) reported that so far as academic subjects were concerned, young naval ratings learn just as well from a teaching machine as from a teacher. Evidence was also obtained in that work which showed that the learning rate, using programmed methods, was much higher than the learning rate achieved by teachers. This finding was confirmed by the other two studies. McKnight (20) compared programmed instruction with formal class methods in the teaching of trigonometry to R.A.F. personnel. Again it was shown that considerable time saving was achieved when programmed learning was used. The third of these three papers, describes a study by Cavanagh, Thornton and Morgan (21) who compared programmed learning with conventional methods in teaching B.E.A, personnel. Once again evidence is put forward to show that the time taken for the Group learning from a programme was about half that for the Group learning from conventional teaching methods. Both Groups, however,

learned to the same extent.

Smith (22), investigating the teaching of statistics in the United States Air Force Academy concludes that neither programmed learning nor conventional teaching produce the better learning but that the time required to achieve the learning is much reduced if programmed learning is used.

Rather more encouraging results have been obtained by Green, Weiss and Nice (23) who have reported that programmed learning is an effective technique in the teaching of medical parasitology. Unlike the research reviewed above this work claims a significantly higher test performance for groups given programmed learning compared to those given conventional teaching. The improvement in performance was manifest in two ways which depended on the ability of the students. Those of high ability who would have scored well on the test in any case achieved their usual performance with much less time spent on study. Those students of lower ability achieved test scores of a level higher than would normally have been the case and did so with less study time than might have been expected. It would seem from these results that the lack of significant difference between pre- and post-tests in the other researches reported earlier, might very well be due to the nature of the group comprising the experimental class. If the groups consisted of only bright students or if the number of less able students was quite small

no gain would be detected.

It might well be the case, however, that this gain reported by Green et al is simply a case of regression towards the mean, since no effort appears to have been made to allow for regression effects.

More recently Elton et al (24) in an experiment to compare a conventional lecture with a programmed lecture selected two topics considered to be of equal difficulty, namely polymerisation and metallic bonding. The polymerisation was taught by giving a normal lecture, while the metallic bonding was taught by programmed lecture.

The programmed lecture consisted of material containing a number of questions, which the students were required to answer during the course of the lecture by putting their replies down on a specially prepared sheet. The correct answers were then immediately given before proceeding to the next question. The mean gain from pre-test to post-test was 45% for the programmed lecture and 12% for the conventional lecture. In spite of this evidence in favour of a programmed approach Elton's students apparently prefer lectures. This, Elton claims, may be due to the lecturers' enthusiasm for this form of teaching, transferring to the students.

Experimental criteria for comparative studies

The validity of a number of these experiments must, however, be questioned, since in many cases the

criteria are not clearly stated or are even lacking
altogether.    In any experiment, those variables which
are not controlled are those which are being measured.
Thus in order to prove differences in effectiveness
between two different methods of teaching, certain
criteria have to be met.    These are:

(i)    content: Two presentations involving two
                different teaching techniques must
                include the same concepts examples
                and illustrations.

(ii) extraneous material: Any material which might
                possibly interfere with efficient
                learning of the relevant material
                should be excluded.

(iii)           Both methods must be equally good.
                It is pointless to compare a good
                lecturer with a badly written programme
                or vice-versa.

(iv)            The time allowed to each method should
                be the same.

(v)             When designing the test care must be
                taken to ensure that the type of
                response mode selected is advantageous
                to neither group.

In the research, already referred to, by Green
et al the criterion of identical content was not met,
since the group of students, who did not have the
programme available to them had to search for the
relevant information.    Support for Green's findings
is supplied by Hughes and McNamara (25), who compared
a programmed text with lecture-discussion, in the teaching
of computer programming.    Here again, however, the
design of the experiment was seriously at fault since
the controls were inadequate.    Those students who were
receiving programmed instruction were allowed to take

the programmes home with them and were allowed to study them at will.   No check was made on the amount of time spent on the programme outside class time.   No mention is made, in this report whether the two techniques covered the same material or not.

Another poorly designed piece of research is reported by Ferster and Sapon (26) who compare programming with conventional methods in the teaching of German. This research fails to meet several of the major criteria. Firstly, their design does not include a control group, the authors simply basing their results on achievement scores on a test designed for the experimental group. Secondly the material covered was not the same as would normally have been the case with the students not taking part, nor was the test itself the test taken by other students in the same course but not involved in the experiment.

This literature survey has attempted to show that, as far as the effectiveness of programmed learning is concerned, many questions remain unanswered.   This would seem particularly true with comparative studies where results are to say the least, conflicting and the soundness of experimental design frequently open to question.   In view of these shortcomings, there is an obvious need for further, well designed research in this area of study.

## CHAPTER 2

'It is a capital mistake to
theorise before one has the
data.'
                    Conan Doyle.

During the past twenty-five years, science education
has undergone a major revolution, which has not only altered
the material content of courses, but also influenced both
the objectives and methods of science teaching.
Unfortunately, the enthusiasm of some educationists
(mainly in positions of high administrative authority) for
change, has led, as pointed out by Cronbach (18), to the
large scale adoption of methods and philosophies as yet
unvalidated by well designed research.

While criticism of this sort is not wholly justified
for programmed learning, it is nevertheless the case, that
much of the comparative research, which has been carried
out to date is of doubtful validity, due largely to
questionable experimental design.

Clearly there is a case for well designed research
comparing chemistry teaching by programme with chemistry
teaching by lecture.  In order to contribute to this need
the present research was undertaken involving two
investigations of differing experimental design.  In the
first of these experiments an attempt has been made to
discover the relative effectiveness of programmed learning
and conventional lecturing in the teaching of infrared and
ultraviolet spectroscopy to undergraduate students

majoring in chemistry.    In the second, the teaching of
ionic equilibria to undergraduates taking chemistry as
a service course, has been investigated using both
programmed learning and lecturing.    These courses, which
are commonly given in the first year of most Scottish
universities, were investigated in the Department of
Chemistry, Heriot-Watt University, Edinburgh, without
whose co-operation this study would not have been possible.

## Aims and scope of the research

### The sample

Since it is impossible, in the majority of
situations, to measure a particular attribute for a whole
population, the normal procedure is to select a sample
which will be representative of the population, as a
whole.

When carrying out an experiment on undergraduates
the obviously correct procedure would be to select, at
random, a sample of students who represented the whole
population of undergraduates reading chemistry at first
year level in Scottish Universities.    The scope and
limitations of this present research necessitated confining
investigations to one university department.    A fairly
serious difficulty therefore, exists within the framework
of any experiment, designed to function with such a sample,
namely that of generalisation.    How far can the results
of such experiments be generalised beyond the confines
of Heriot-Watt University?    It may very well be the case

that undergraduates reading chemistry at Heriot-Watt
are in fact representative of undergraduates generally
in that their attitudes to different forms of teaching
are largely the same.   Hogg (17) did not report any
marked differences among undergraduates from several
tertiary education centres, in attitudes to programmed
learning.   It is possible that Heriot-Watt students
differ from other undergraduates in an intellectual way
but this should not affect the generalisation of results,
since there is no evidence from research reviewed in
Chapter 1 that success with any particular teaching
method is a function of the intellectual ability of the
learner.   In any case, such intellectual differences
that might exist, will, as a result of selection for places,
be rather small.

In socio-economic terms there is certainly no
evidence that Heriot-Watt students are not representative
of the student body generally (27).   It would seem,
therefore, that even without a random sample, taken from,
say all Scottish Universities, generalisation of the
results beyond the confines of Heriot-Watt University is
probably justified.   It might be argued, however, that
since every member of the population did not have an equal
opportunity of being selected for the sample, the sample
is in this sense biased, and to that extent generalisation
regarding the population could be less accurate than
might be hoped.

## Internal and external validity

When designing an experiment it is necessary to
consider the principles of internal and external validity
(28). A piece of research which is internally valid is
a study in which the measured outcomes can be shown to be a
function of a particular independent variable and not a
function of some other variable which was not taken account
of in the experiment. A piece of research which is
externally valid is a study whose results or measured
outcomes are generally applicable to other similar situations.

In the carrying out of an experiment the control
of certain variables improves or ensures the internal
validity but at the same time limits the external validity.
In other words, as the circumstances under which the
investigation is carried out are controlled, the probability
is that the measured outcomes are the result of changes
in the independent variable only, that is, the experiment
is internally valid. Unfortunately the more control is
exercised, the less likely is it that the experimental
conclusions will apply to a situation other than the
experiment itself, since in the real world, the experimental
controls are absent. However, unless some action is taken,
in an experiment to control variables and provide a degree
of internal validity it may never be discovered what
caused the observed effects. Consequently external
validity is of little value without internal validity to
provide a measure of confidence in the results. An
experiment which is internally valid may also be externally

valid.   One which is not internally valid can never be
externally so.

Before it may be claimed that the observed effects
are a function of the independent variable the internal
validity of the design must be checked.   There are a
number of factors which might well be responsible for
the observed effects.   These are:

1.   Contemporary history:

Occasionally experimental subjects experience an
event either within or without the experimental situation
which affects the dependent variable scores.   Students
taking part in an experiment to compare two teaching
methods might possibly see a spectacular television
programme (dealing with the same topic as the experiment)
prior to taking the post test.   This uncontrolled variable
then, has confounded the independent variable since the
observed effects are attributable to two or more variables.

2.   Maturation:

Psychological and physical processes within the
experimental subjects may undergo changes between the
treatment and the post-test.   Students may perform less
well than might be expected, on a post-test simply because
they are fatigued by the process or simply bored by it.
The effect might be in the other direction - learning
might have taken place between the pre- and post-tests in
ways not taken account of in the teaching process.

3.   The Pre-Test:

There is always the possibility that tests themselves

provide a learning experience. Even in a situation where no experimental 'treatment' is used a pre-test may cause subjects to alter their responses to a post-test.

## 4. The test instruments:

If the post-test is more difficult than the pre-test or if these two tests are scored by different people these factors, rather than the experimental treatment, could produce the observed effects. Even if the same experimenter scores both the pre-test and the post-test, his judgment may not be constant from one test to the next simply because he becomes more experienced or fatigued.

## 5. Regression:

In some areas of educational research, experimental subjects are selected on the basis of their extreme nature on some variable. When this sort of selection is employed the effect of regression may be mistaken for the effect of the treatment. If, for example, only those students who do well on a pre-test are tested a second time the mean score of this selected group will move towards the population mean (i.e. downwards whether or not any treatment is applied between the first and second occasions of testing, giving the entirely false impression, in an experiment, that the change is due to the treatment. In a similar situation the mean of a low extreme group will be inevitably higher on a second occasion of testing. Regression towards the mean is a

phenomenon which is the result of random imperfections in the tests. If the correlation between pre- and post-tests is less than perfect (usually the case) each experimental subject will obtain similar but not identical scores on each test, their scores varying within a given range. In other words, there will be a tendency for extreme scores on a pre-test to tend towards the mean on any subsequent test since the chance factors, which contribute more to extreme scores than to average scores, are less likely to operate in the second case.

6. Differential selection:

Any difference between a control group and an experimental group detected by a post-test may be due to differences between these groups prior to the administration of any treatment.

7. Experimental mortality:

There is always the possibility in an experimental situation that some subjects may drop out after the administration of the pre-test and before the administration of the post-test. If the drop-out is random across both groups then the internal validity of the experiment is unaffected. If, however, only those subjects who did badly on the pre-test drop out then any gain detected by the post-test will be artificially high.

There are many factors which affect the external validity, or generalisability of an experiment. These are:

1. Interaction of selection biases and the treatment:

The characteristics of subjects selected to take
part in an experiment influence greatly the generalisability
of the results.   A random sample of first year pupils
from school X may not represent all first year secondary
pupils.   The I.Q., or socio-economic status, of the
selected pupils may cause the experimental treatment for
them to be more effective than for other first year pupils.

2.  The effect of pre-testing:

The administration of a pre-test can affect the
generalisability of findings.   The pre-test can, for
example, "sharpen" an experimental subject's sensitivity
to the treatment.   It could alert them to problems or
issues which they otherwise might not have noticed.   As a
result of this, these experimental subjects are no longer
representative of the unpre-tested population from which
they came.   Solomon's design may be used to offset the
effect of pre-testing by regarding the pre-test as
another treatment.   The post-test scores can then be
treated in a 2 x 2 design thus:-

|  | Teaching Method A | Teaching Method B |
|---|---|---|
| Pre-tested Group |  |  |
| Unpre-tested Group |  |  |

3.  The effect of experimental treatment:

The administration of the experimental treatment
itself can often affect the outcome depending on how it

is administered. If there are present any observers
or any unusual or sophisticated equipment this makes
experimental subjects more aware of the fact that they
are involved in an experiment and may thus alter their
behaviour (the Hawthorne effect). Thus it cannot be
claimed unreservedly that any effect detected by the
post-test would apply to subjects exposed to the treatment
in a non-experimental situation.

## The hypothesis

Having argued for cautiously accepting that the
sample of students chosen for research purposes represents
the population, namely all first year undergraduates
reading chemistry at Scottish universities, the hypothesis
to be tested must now be stated. The problem under
investigation necessitates formulating an hypothesis
stating that there will be no significant difference in
performance on a post-test, between initially equivalent
groups taught by different methods: in other words,
that the two methods (in Fisher's parlace 'treatments')
are equally effective. The hypothesis has to be a null
hypothesis since it is effectually saying that the two
groups compared come from the same population. Any
experiment which could be carried out to test such an
hypothesis must be designed to compare the mean scores
of two groups, the members of which have been randomly
assigned to the two groups from the original sample,
which itself, it is hoped, is representative of the

population. If a difference in mean is obtained the probability that it is due to sampling error (that is, to chance) can readily be assessed using analysis of variance techniques (29).

Designing the experiment - control of variables.

Having set up the null hypothesis an experiment has next to be designed in which performance on some test of chemical ability is examined in relation to teaching techniques. The independent variable in such an experiment would be the teaching method, the dependent variable, scores on a chemistry test, the aim therefore being to test the hypothesis that changes in the independent variable have no effect on the dependent variable.

It is, as stated earlier, clear that performance in any test of chemistry (or any other discipline) could well be related to some variable other than the method by which the material concerned is taught. Such extraneous variables cannot be ignored. Any experiment must therefore be designed so that the effect upon the dependent variable is attributable to the independent variable, and not to some unidentified variable or variables.

One method eliminating this confounding of the effects of independent and dependent variables is to control by randomisation. If individuals are assigned to either of two groups (experimental and control) by random number selection it is reasonable to assume that

the two groups would be comparable on most variables.
Within this present study, therefore, two randomly
selected groups will be initially comparable on the
dependent variable of chemical ability and also on
extraneous variables such as intelligence.   Thus if any
difference is detected between the experimental and control
groups on a post-test of chemical ability, it is fair to
assume that such a difference is unrelated to the
intelligence variable.   Intelligence is not, however, the
only extraneous variable which might have an effect on the
dependent variable, chemical ability.   Others might be,
for example, age, attitudes, sex, social factors, perhaps
even hereditary factors, all of which are allowed for in
a control by randomisation design.

In addition to randomisation, it is possible to
make groups comparable on some extraneous variable by
making the groups as homogeneous as possible on that
variable.   If, for example, I.Q. were thought to be a
factor which would affect an experimental outcome, in such
a way that any effect due to the experimental treatment
would be obscured, the control and experimental groups
could be made homogeneous with respect to I.Q., and thus
any effect due to I.Q. would be allowed for.   However,
the effects of other unrelated extraneous variables such
as age, sex, motivation and so on, would still remain;
and there is an obvious limitation on the number of
extraneous variables that can be treated in this way.

First year students are however a highly selected

group with narrow age and I.Q. ranges, and therefore
two randomly assigned groups could very well be
reasonably homogeneous on those two variables also.
Homogeniety of groups, will, as pointed out previously,
increase the likelihood that the groups are comparable
(at least on the variables considered).  At the same time,
however, it also reduces the generalisability of the
results since populations are seldom homogeneous on many
variables.  In this particular case, however, wide
generalisation is not required since the aim of this
research is to compare programmed learning with conventional
lecturing for undergraduate students only.

The type of generalisation which would be unjustified
is generalisation to an age range of 12 to 18 years
regarding the relative effectiveness of the two teaching
methods employed.

It is not possible (no matter how much case is taken
to ensure sample representativeness and group comparability)
to ensure that all error is eliminated from data.
Statistical techniques are therefore required to estimate
the probability that an observed difference between two
means is due to sampling errors rather than to real
differences in the independent variable.

A classical experimental design such as randomised
groups where as many variables as possible are held
constant while one is deliberately changed has certain
limitations.  The principal difficulty lies in the fact
that there may be a relationship between the experimental

variable under examination and some other variable (held constant by the randomisation procedure) the interaction of which produces an outcome in excess of that which would have resulted from either of the two variables alone.

It might be the case, for example, that a piece of research, based on a random group design, showed that physics teachers who used an 'open-ended' style of teaching were more effective in the teaching of their subject than physics teachers who adopted an authoritarian approach. This, in itself, is a useful finding but could have been more important if some other variable had been investigated at the same time. If, for example, it had been shown that the 'open-ended' style physics teachers were more successful than 'authoritarian' style teachers, with pupils of say I.Q. 115 and above - and the authoritarian teachers more effective with pupils of I.Q. less than 115, then such a result would be much more important than that produced by the random group design since it detects an interaction between teaching styles and pupil I.Q.

An experimental design which allows for the manipulation and checking the effects of two or more variables simultaneously rather than in separate experiments is a factorial design. Such a design is an advance on two separate random group designs since the interaction (joint) effect of the two variables may be calculated.

In any research investigating the relative effectiveness of two teaching methods, a possible interacting

variable is subject material. A simple 2 x 2
factorial experiment could be designed thus:

|  | Lecturing | Programmed Learning |
|---|---|---|
| Subject Material I | Group 1 | Group 3 |
| Subject Material II | Group 2 | Group 4 |

Such a design could obviously be extended to a $\vec{2}$ x↓n
design.

Students would be allocated at random to each cell
and exposed to the variables, for example Group 1 would
be taught subject material I by lecturing, Group 3 taught
the same material by programmed learning and so on.
Finally a post-test would be given the results of which
would provide data from which three questions could be
answered

(1)   What is the main (independent) effect of lecturing
      on the post-test scores ?

(2)   What is the main effect of subject material on
      post-test scores ?

(3)   What is the interaction effect, if any, of teaching
      method and subject material on post-test scores ?

The 2 x 2 factorial design described above may
sometimes require modification depending on circumstances.
If, for example, random assignment of students to groups
were not possible and intact classes had to be used or if
the number of available students were small then a

"counterbalanced" or "cross-over" design might be
employed.   In such a design groups of students
(randomly assigned where possible) are allocated to each
teaching method in turn.   In a situation where there
are two groups and two teaching methods, group A would
be exposed to Method X and group B exposed to Method Y
on the first of two occasions.   On a second subsequent
occasion the teaching situations are reversed producing
a design known as a 2 x 2 Latin square.

|         | Time 1 | Time 2 |
|---------|--------|--------|
| Group A |   X    |   Y    |
| Group B |   Y    |   X    |

The 2 x 2 factorial design and the 2 x 2 Latin
square are similar in some respects.   In the above table
the row and column sums of squares are the same for
both designs while the row and column "interaction" in
the 2 x 2 factorial design is identical with "treatment
effect" in the 2 x 2 Latin square.

Experiment 1

The undergraduates selected for this experiment
were intending honours students in their first year of
study during the academic year 1972-1973.   The class was
divided randomly into two groups which were designated A
and B.   Group A was taught infrared spectroscopy (i.r.)
by programmed learning (30) and ultraviolet spectroscopy
(u.v.) by lecturing.   By incorporating a change in
material content and reversing the order for group B

a 2 x 2 Latin square design was possible.

While group A was given a one hour lecture on ultraviolet spectroscopy by Dr I. Soutar group B worked on the programme in a nearby tutorial room.  When the lecture was complete group B were asked to stop working and indicate how much of the programme they had completed.

The following day the teaching positions were reversed group B attending a one hour lecture on infrared spectroscopy (again given by Dr I. Soutar) while group A worked the programme.  When this second lecture was complete group A students were asked to stop working and indicate what proportion of the programme they had completed. Finally both groups were asked to state their preference either for lecturing or programmed learning.

These preliminary investigations showed that in both groups 90% of students preferred lecturing to programmed learning and surprisingly, (in view of the research evidence reviewed in Chapter 1) no student completed the programme, the majority attaining between 50% and 75% completion.

Both groups were subsequently tested by means of two thirty minute tests, one testing infrared the other ultraviolet spectroscopy.  (Appendix I)

The data are summarised in Table I which shows total scores with means in brackets.  The complete table of results is shown in Appendix II.

## TABLE I

Summary of scores obtained in Experiment I

### Teaching Methods

|  | Lecture | Programme | Total |
|---|---|---|---|
|  | Ultraviolet scores | infrared scores |  |
| Group A (9Ss) | 480 (53.3) | 619 (68.8) | 1099 (61.1) $T_A$ |
|  | infrared scores | ultraviolet scores |  |
| Group B (9Ss) | 625 (69.4) | 468 (52.0) | 1093 (60.7) $T_B$ |
| Total | 1105 (61.4) $T_L$ | 1087 (60.4) $T_P$ | 2192 |

(Means in brackets)

This type of experimental design is based on the
2 x 2 Latin square. This means that in addition to a
direct comparison of one teaching method with another there
is also the opportunity to consider such sources of
variation as content material, students within groups,
and interaction within groups. The results require to
be analysed in two parts (at least initially) two-way
analysis of variance being applied to Groups A and B
separately and finally pooled.

## Analysis of variance - sums of squares

When computing the sum of squares for two sets
of scores $x_1$ to $x_n$ and $y_1$ to $y_m$ the expression used is:

$$ss = \frac{(\Sigma x)^2}{n} + \frac{(\Sigma y)^2}{m} - \frac{(\Sigma x + \Sigma y)^2}{n + m}$$

In a situation where $n = m$ this expression
simplifies to:

$$ss = \frac{(\Sigma x - \Sigma y)^2}{2n}$$

The sum of squares between teaching methods
(Lectures v. Programme) will be given by

$$\frac{(T_L - T_P)^2}{36} = \frac{(1105 - 1087)^2}{36} = 9 \quad d.f. = 1$$

Similarly the sum of squares between groups
(Group A v. Group B) will be given by:

$$\frac{(T_A - T_B)^2}{36} = \frac{(1099 - 1093)^2}{36} = 1 \quad d.f. = 1$$

The sum of squares for subject material (i.r. v. u.v.)
will be given by:

$$\frac{(T_{i.r.} - T_{u.v.})^2}{36} = \frac{(1244 - 948)^2}{36} = 2433.778 \quad d.f. = 1$$

Explanation of statistical treatment

A two-way analysis of variance was carried out
on each group separately.  These two analyses were pooled
to give a within groups analysis of variance.  Consider
then the situation when a group is taught one section of
the syllabus (a) by lecturing and another section (b)
by programmed learning.  The sources of variation in a
two-way analysis of variance would be

        Students    d.f. $(n - 1)$ (assuming n students per group)
        Content     (a v b) d.f. $(m - 1)$ (assuming m methods)
        Interaction (S x C) d.f. $(n - 1)(m - 1)$

For both teaching groups A and B the sources of
variation and the degrees of freedom are:

        Students     d.f. = 8
        Content      d.f. = 1
        Interaction  d.f. = 8

Two-way analysis of variance on Group A

1.    $\sum x^2 = 44061 + 28416$      $= 72477.00$

2.    $\dfrac{\sum(x_{i.r.} + x_{i.v.})^2}{2}$      $= 69606.50$

3.    $\dfrac{(\sum x_{i.r.})^2}{9} + \dfrac{(\sum x_{u.v.})^2}{9} = 68173.44$

4.    $\dfrac{(\sum x_{i.r.} + \sum x_{u.v.})^2}{18}$      $= 67100.06$

| Source of Variation | d.f. | Sum of Squares | Mean Square |
|---|---|---|---|
| Content (i.r. v u.v.) 3 - 4 | 1 | 1073.38 | 1073.38 |
| Students 2 - 4 | 8 | 2506.44 | 313.31 |
| Interaction | 8 | 1797.12 | 224.64 |
| Total   1 - 4 | 17 | 5376.94 | |

Check calculations

Since there is only 1 degree of freedom the content sum of squares for group A only will be given by

$$ss = \frac{(619 - 480)^2}{18} = 1073.38$$

The sum of squares for students will be given by

$$\sum\nolimits' \frac{(a + b)^2}{2} = \frac{1}{2}\left\{\sum(a + b)^2 - \frac{(\sum(a + b))^2}{9}\right\}$$

where a = i.r. scores and b = u.v. scores.

$$ss = \frac{1}{2}\left\{139213 - \frac{(1099)^2}{9}\right\}$$

$$= 2506.44$$

The interaction sum of squares will be given by

$$\sum\nolimits' \frac{(a - b)^2}{2} = \frac{1}{2}\left\{\sum(a-b)^2 - \frac{(\sum(a - b))^2}{9}\right\}$$

$$= \tfrac{1}{2} \left( 5741 - \frac{(139)^2}{9} \right)$$

$$= \underline{1797.11}$$

## Two-way analysis of variance on group B

1. $\sum x^2$ $\qquad = \quad 71181.00$

2. $\dfrac{\sum (x_{i.r.} + x_{u.v.})^2}{2} \qquad = \quad 68130.50$

3. $\dfrac{(\sum x_{i.r.})^2}{9} + \dfrac{(\sum x_{u.v.})^2}{9} \qquad = \quad 67738.78$

4. $\dfrac{\sum x_{i.r.} + \sum x_{u.v.}}{18} \qquad = \quad 66369.39$

| Source of Variation | d.f. | Sum of Squares | Mean Square |
|---|---|---|---|
| Content 3 - 4 | 1 | 1369.39 | 1369.39 |
| Students 2 - 4 | 8 | 1761.11 | 220.14 |
| Interaction | 8 | 1681.11 | 210.14 |
| Total 1 - 4 | 17 | 4811.61 | |

## Check calculations

Again since there is only 1 degree of freedom the sum of squares for content is given by

$$ss \quad = \quad \frac{(625 - 468)^2}{18}$$

$$= \quad \underline{1369.39}$$

As previously described the sum of squares for students is given by

$$ss \quad \frac{\sum'(a + b)^2}{2} = \tfrac{1}{2} \left\{ \sum (a + b)^2 - \frac{(\sum (a + b))^2}{9} \right\}$$

$$= \frac{1}{2} \left(136261 - \frac{(1093)^2}{9}\right)$$

$$= \underline{1761.11}$$

The interaction sum of squares will again be given by

$$\frac{\sum'(a-b)^2}{2} = \frac{1}{2}\left\{\sum(a-b)^2 - \frac{(\sum(a-b))^2}{9}\right\}$$

$$= \underline{1681.11}$$

When the two analyses are pooled one degree of freedom is added for the between groups variation and the two degrees of freedom for content divided into one for content and one for method. The pooled analysis of variance table is obtained thus.

| Source of Variation | d.f. | Sum of Squares | Mean Square | F |
|---|---|---|---|---|
| Between Groups | 1 | 1.00 | 1.00 | $\frac{1}{266.72}$ N. |
| Content (i.r. v. u.v.) | 1 | 2433.78 | 2433.78 | $\frac{2433.78}{217.39}=11$ |
| Methods (Lect. v. prog.) | 1 | 9.00 | 9.00 | $\frac{9.00}{217.39}$ N. |
| Students within groups | 16 | 4267.55 | 266.72 | |
| Interaction with groups | 16 | 3478.23 | 217.39 | |
| Total | 35 | 10189.56 | | |

F = 11.21 with degrees of freedom (1, 16) is significant at the 1% level.

Check calculation for total sum of squares:

$$\sum x^2 - \frac{(\sum x)^2}{n} = 143658.00 - 133468.44$$

$$= \underline{10189.56}$$

It is not possible to ensure, in any experiment
such as this, where two groups are compared, that (a)
the sample is fully representative of the population and
(b) the experimental and control groups are equal in all
respects.    Statistical techniques such as those described
above allow estimates to be made of the probability that
observed differences between groups are due to random
errors rather than to real differences in the independent
variable.

In rejecting or accepting a null hypothesis there is
always an element of risk that an incorrect decision will
be made.

Although the level at which the null hypothesis will
be rejected is normally determined beforehand two types
of error are possible, namely rejecting a null hypothesis
which should be accepted (Type I) and accepting a null
hypothesis which should be rejected (Type II).

In rejecting a null hypothesis based on a statistical
test of sample data there is always the possibility that
a Type I error will be made.    At the 5% level the chances
of doing so are quite small;   at the 1% level even less.
Thus the lower the probability level the less chance there
is of making a Type I error.    Unhappily the lowering of
the probability level increases the chance of a Type II
error.

The results of Experiment 1 seem, therefore, to
indicate that, regardless of subject material a slightly
better performance (in terms of means) was obtained by the

group which attended the lectures. The difference, however, is shown by the analysis of variance not to be significant. It may be concluded that, in <u>this</u> situation with <u>this</u> subject material, <u>these</u> lectures and <u>these</u> programmes, programmed learning and conventional teaching are equally effective teaching methods.

This is a slightly surprising result, since the students indicated, not only dislike for the programme (90% preferring lectures) but also an inability to cover the material at a rate comparable to that of the lecture. At first sight it might appear that students working on the programme work slowly. It seems unlikely that this is the case, a more credible explanation being associated with the larger number of illustrative example s in the programme than in the lecture. The student who diligently works through all these examples will, of necessity, cover much more material, much of it repetitive, than the student attending the lecture. It may be that students require some 'training' in the use of programmes or at least an assurance that if they can cope with say, two illustrative examples it is not necessary to work through all. Thus a programme would be much more efficiently used and, indeed, the results of enquiries such as this might easily show programmed learning to be the more effective teaching method.

As pointed out in Chapter I, the well-designed comparative experiment of this kind must endeavour to ensure that the depth of treatment is the same in both

teaching methods.   In the ultraviolet test question 2
covered material which was thoroughly dealt with by the
programme but treated cursorily by the lecture.   For
this particular question, at least, those students who
studied the programme produced better answers.   The
type of question, of which question 2 is an example,
(See Appendix I) involves low level cognitive objectives
such as recall and comprehension.   The questions which
involved problem solving such as question 6 were, in
general, better answered by those students who had attended
the lecture.   This, of course, probably reflects the
fact that student performance is influenced to a large
extent by emphasis, on the part of the teacher.

Perhaps the most interesting outcome of this experiment
is the significant difference in performance between the
two subject areas regardless of group or teaching method.
The results clearly show a much poorer performance in the
test on ultraviolet spectroscopy.

There are several possible reasons for this difference.
Firstly, it might very well be the case that the test on
infrared spectroscopy was simply easier than that for
ultraviolet.   It is difficult to avoid this type of
situation unless objective testing is used, when tests
previously tailored to be of equal average facility value
can be employed.   Secondly the order in which the subject
material was taught placed ultraviolet spectroscopy
before infrared and this might possibly have had some
influence on the scores.   Thirdly ultraviolet spectroscopy

may be conceptually more difficult than infrared. Recent
work in the University of Glasgow Chemistry Department (31)
has indicated that, at school level, certain topics
within the chemistry syllabus are found by the pupils to
be more difficult than others.  It might very well be
the case, that with certain students at least, if not all,
some topics in chemistry are conceptually difficult if
taught in the early stages of the course and might be
better understood if left until a later year.  It might
even be the case that Piaget's ideas of stage developments
among children apply also to young adults.  Indeed, at
what age does the intellectual development of an
individual stop ?  If intelligence, as measured by
intelligence tests, is directly related to intellectual
capacity, it would seem that the peak of intellectual
power comes, according to Wechsler (32), at about the age
of thirty.  There would seem to be therefore, a case
for the extension of Piaget's ideas to learning situations
beyond the school level.

The results of this particular experiment would
appear to indicate that, while students prefer to be given
lectures rather than programmes, and while they cover
much less material when they do use a programme, the
programmed method is as effective in teaching certain
aspects of spectroscopy as the more conventional lecture
technique.

The experiment also showed that regardless of the

teaching technique involved not all areas within the subject are equally well understood.

## Experiment 2

The experimental design employed in this part of the investigation involved the use of simple group randomisation. The sample of students was considerably larger than that employed in Experiment 1, the majority taking chemistry purely as a service course. The sample included students reading pharmacy, biochemistry, brewing, microbiology and chemical engineering, a total of 137 students at the start of the experiment. The sample was divided into two groups designated A and B using a list of random numbers provided by a computer, the programme for which is shown in Appendix III. A test of general chemical ability (Appendix IV) was administered to the whole sample, the scores on which were used as pre-test data. Group A were then given lectures on ionic equilibria, while Group B worked through the programme (33). The experimental conditions applying were those described for Experiment 1. A post-test on ionic equilibria (Appendix V) was administered to the whole sample which had suffered a mortality of approximately 14% reducing the sample size to 96. The effect of this mortality on the random nature of the two groups was assessed by calculating the mean pre-test scores for the drop-outs in each group. The means were not significantly different and consequently the mortality suffered by the experiment did not affect the random nature

of the two groups.

Many experimental designs involve the experimental
control of a concomitant variable, or variables, while
measures are made of the effect of some independent
experimental variable. It is not always possible,
however, to control a concomitant variable for reasons
of administrative inconvenience to those in charge of
schools or university departments and consequently
control methods such as random assignment and pair matching
cannot be used. It is sometimes the case that variability
in a concomitant variable it not obvious to the researcher
until after the experiment has been carried out. In a
comparison of teaching techniques, for example, the amount
of time students devote to private study might well be an
important variable, easy to measure but impossible to
control experimentally. When control of such a variable
is impossible, or at best difficult to ensure experimentally,
control by statistical technique is possible.

Statistical control of a variable is exercised by
making observations of a concomitant variable and
"adjusting" the criterion means for the various treatment
groups. Such a technique is known as analysis of
covariance (34).

The experiment here described is a simple randomised
group design consisting of two treatments, lecturing and
programmed learning. Scores are available for a post-test
(the criterion) and a pre-test (the concomitant variable).
The post-test relates directly to the subject matter

taught by the two teaching methods, while the pre-test
is a test of general chemical ability which is not a
controlled variable within the design of the experiment,
since scores on a test of general chemical ability are a
function not only of understanding (which is controlled
by randomisation) but also of amount of time students
devote to private study, which is not controlled by
randomisation (27).    In other words it is possible to
equate statistically the two groups on chemical ability
so that differences on this extraneous variable would not
differentially affect the dependent variable scores.

Covariance is defined as the product of the deviations
of scores on a variable X from its mean $\bar{X}$ and deviations
of scores on a variable Y from its mean $\bar{Y}$.    If a number of
experimental subjects have a series of X scores, say
intelligence and a series of Y scores, say reading scores,
then the experimental results consist of a series of
ordered pairs of scores.    If the deviations $x$ and $y$ are
listed (where $x = (X - \bar{X})$ and $y = (Y - \bar{Y})$ another set of
ordered pairs is obtained.    If the X and Y columns are
then multiplied and finally summed a measure, (the sum of
products) is obtained which is analogous to the sum of
squares in analysis of variance.    Whereas analysis of
variance is concerned with sums of squares and variances,
analysis of covariance is concerned with sums of products
and covariances in addition to sums of squares.

The essence of the process in analysis of covariance

is that the significance of any difference between the
Y means after adjustment may be readily tested.    The
adjustment effectually removes from the Y sum of squares
that part which is due to the relationship between X and Y.
The higher the correlation between X and Y the more
effective is the analysis of covariance.    The final
adjusted analysis table gives an adjusted total, and
sums of squares for between and within groups.    The
variances (mean squares) and the F ratio are calculated
from these adjusted measures.

If a difference is detected between the post-test
adjusted means it may be reasonably claimed that it is
not due to any difference in the factor measured by the
pre-test, but instead to chance or, if significant, to
differing effectiveness in the two teaching methods.
There is also the possibility that a difference in
adjusted means might be due to some extraneous variable
if control by randomisation had been less effective than
hoped.

Effectually what will be achieved through statistical
control is a level of precision in the design of the
experiment comparable to that which might have been
expected had the Groups been matched according to the
variables measured by the pre-test.

The data obtained in Experiment 2 are summarised in
Table II which shows total scores with means in brackets.
The complete table of results is shown in Appendix VI.

## TABLE II

Summary of scores obtained in Experiment 2

|  | Pre-test (x) | Post-test (y) |
|---|---|---|
| Group A (48 ss) | 2612 (54.4) | 3159 (65.8) |
| Group B (48 ss) | 2558 (53.3) | 3325 (69.3) |

## Analysis of covariance - sums of squares

### The pre-test (x scores)

1. $\sum x_A^2 + \sum x_B^2$ = 295712.00

2. $\dfrac{\left(\sum x_A\right)^2}{48} + \dfrac{\left(\sum x_B\right)^2}{48}$ = 278456.416

3. $\dfrac{\left(\sum x_A + \sum x_B\right)^2}{96}$ = 278426.041

### The post-test (y scores)

1. $\sum y_A^2 + \sum y_B^2$ = 475919.000

2. $\dfrac{\left(\sum y_A\right)^2}{48} + \dfrac{\left(\sum y_B\right)^2}{48}$ = 438227.207

3. $\dfrac{\left(\sum y_A + \sum y_B\right)^2}{96}$ = 437940.161

### Products

1. $\sum x_A y_A + \sum x_B y_B$ = 363768.000

2. $\dfrac{\sum x_A \sum y_A}{48} + \dfrac{\sum x_B \sum y_B}{48} = 349097.041$

3. $\dfrac{(\sum x_A + \sum x_B)(\sum y_A + \sum y_B)}{96} = 349190.41$

| Source of Variation | d.f. | Sum of Squares | | |
| --- | --- | --- | --- | --- |
| | | x | y | xy |
| Between Groups 2. - 3. | 1 | 30.40 | 287.05 | -93.38 |
| Within Groups 1. - 2. | 94 | 17255.58 | 37691.79 | 14670.96 |
| Total 1. - 3. | 95 | 17285.98 | 37978.84 | 14577.58 |

From the above table it can be seen that there is no difference between groups on the pre-test x nor on the post-test y, since in both cases $F < 1$.

Pre-test

$$F = \dfrac{\dfrac{30.4}{1}}{\dfrac{17255.58}{94}} < 1$$

Post-test

$$F = \dfrac{\dfrac{287.05}{1}}{\dfrac{37691.79}{94}} < 1$$

Clearly a non-significant value for F in the adjusted analysis is to be expected.

Again from the above table, it can be seen that the total
correlation coefficient between the pre- and post-test
$(r_{xy_{(T)}})$ is negative. The group means for pre- and
post-tests were plotted and a graph confirming this was
obtained (Figure I).

Adjustment of $Y_{TOT}$.

$$y'_{TOT} = y_{TOT} - \frac{xy^2_{TOT}}{x_{TOT}} = \sum'y^2_{TOT}(1 - r^2_{xy(TOT)})$$

$$= \underline{25685.29}$$

Adjustment of $Y_W$

$$y'_W = y_W - \frac{xy^2_W}{xW} = \sum'y^2_W(1 - r^2_{xy(w)})$$

$$= \underline{25218.32}$$

Final adjusted analysis

| Source of Variation | d.f. | Sum of Squares | Mean Sq. | F |
|---|---|---|---|---|
| Between Groups | 1 | 466.97 | 466.97 | F =1.72 (1, 93) |
| Within Groups | 93 | 25218.32 | 271.16 | |
| Total | 94 | 25685.29 | | |

F is not significant at the 5% level as predicted.

The error variance before adjustment of criterion
score is approximately 401. In the adjusted analysis table
this has been reduced to 271 an increase in precision of
about 30%.

The F ratio provides a test of the null hypothesis
that the adjusted means are identical.

The analysis of covariance indicates no significant

FIG 1
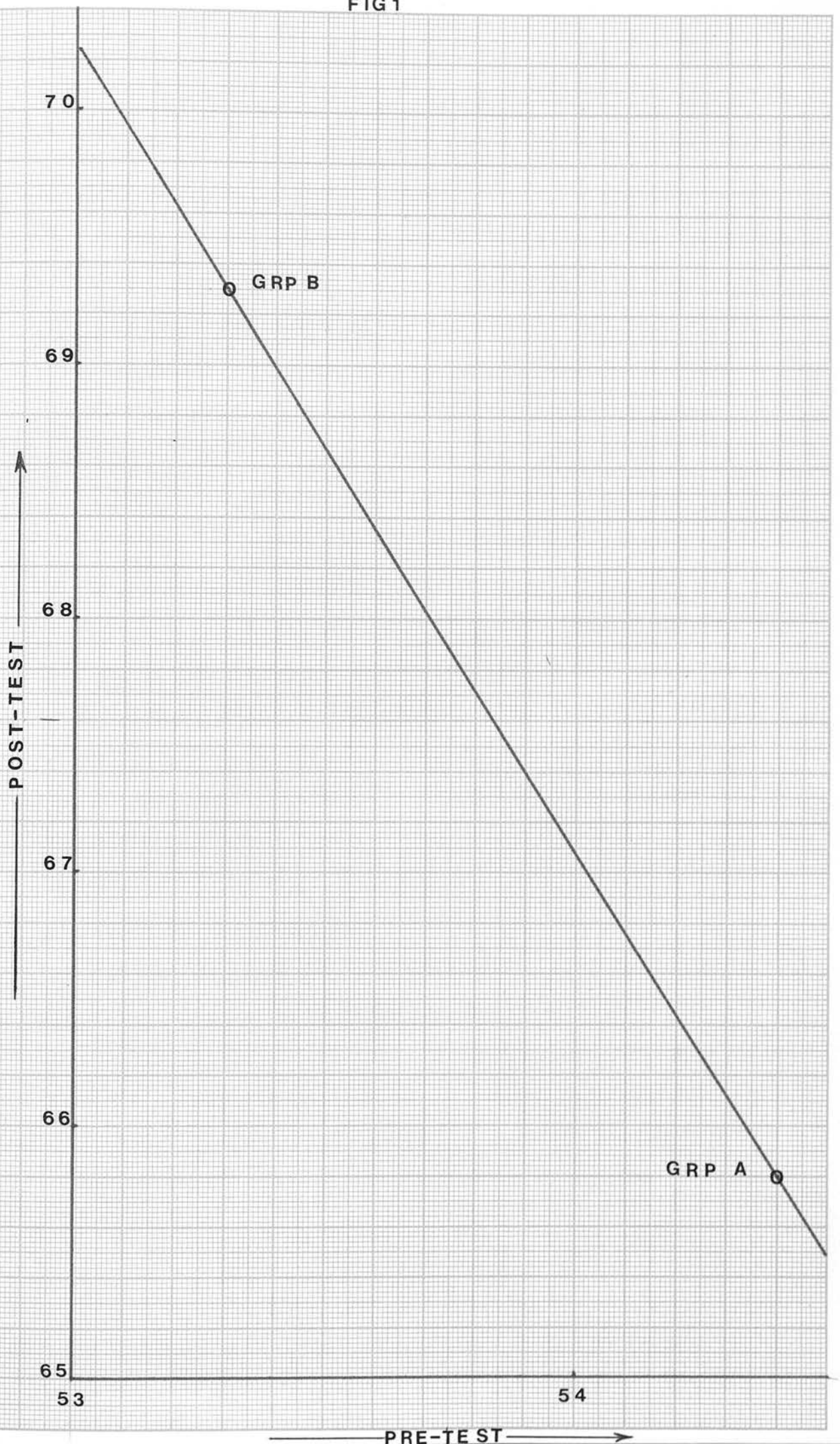
difference between the adjusted means.  Whether this
implies that the two teaching methods are equally
effective depends, in some measure, on how the two
treatment groups were selected.  It is possible, for
example, that the two treatment groups were in fact
different at the start of the experiment and that these
initial differences compensate for differences in the
treatments making the adjusted criterion means equal
(or nearly so) after the experiment.

Since the members of each treatment group in
experiment 2 have been randomly assigned (from a sample
which, it has been argued, is representative of the
population of first year undergraduates taking chemistry)
and since the analysis of the unadjusted data supports
the claim of group comparability, then it may reasonably
be argued that the two treatments (namely lecturing and
programmed learning) used in Experiment 2 were equally
effective in teaching first year undergraduates the
chemistry of ionic equilibria.

## Analysis of the post-test

Since the results of Experiment 2 seem to indicate
that the post-test, as a whole, was unable to detect
any significant difference between treatments it was
decided to analyse the post-test scores question by
question.  It is not possible to compare the means for
each treatment group on each question using the 't' test

$$t = \frac{\bar{d}}{\sqrt{\dfrac{\Sigma'x^2 + \Sigma'y^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

since the number of marks allocated to any one question
(maximum 7) could not possibly be normally distributed.

It was considered useful, however, to arrange numbers
of students in a scores times treatments matrix and
calculate chi-square for each question.   These data can
be seen in Tables III to XI.

TABLE III
Question 2
Scores

|  | 0 | 1 | 2 | 3 | Total | Mean Score |
|---|---|---|---|---|---|---|
| Lecture | 0 | 4 | 3 | 41 | 48 | 2.77 |
| Programme | 2 | 3 | 5 | 38 | 48 | 2.56 |
| Total | 2 | 7 | 8 | 79 | 96 | |

$x^2 = 2.65$ d.f. $= 3$ N.S. at 5% level

TABLE IV
Question 3
Scores

|  | 0 | 1 | Total | Mean Score |
|---|---|---|---|---|
| Lecture | 4 | 44 | 48 | 0.93 |
| Programme | 3 | 45 | 48 | 0.93 |
| Total | 7 | 89 | 96 | |

$x^2 < 1$   d.f. $= 1$   N.S. at 5% level

TABLE V
Question 4
Scores

|  | 0 | 1 | 2 | 3 | 4 | 5 | Total | Mean Score |
|---|---|---|---|---|---|---|---|---|
| Lecture | 9 | 8 | 3 | 8 | 9 | 11 | 48 | 2.69 |
| Programme | 7 | 9 | 5 | 9 | 13 | 5 | 48 | 2.56 |
| Total | 16 | 16 | 8 | 17 | 22 | 16 | 96 | |

$x^2 < 9$ d.f. $= 5$ N.S. at 5% level

## TABLE VI
### Question 5
#### Scores

|  | 0 | 1 | 2 | 3 | 4 | Total | Mean Score |
|---|---|---|---|---|---|---|---|
| Lecture | 3 | 0 | 1 | 7 | 37 | 48 | 3.6 |
| Programme | 3 | 3 | 0 | 6 | 36 | 48 | 3.4 |
| Total | 6 | 3 | 1 | 13 | 73 | 96 |  |

$x^2 < 9$   d.f. = 4   N.S. at 5% level

## TABLE VII
### Question 6
#### Scores

|  | 0 | 1 | 2 | 3 | 4 | Total | Mean Score |
|---|---|---|---|---|---|---|---|
| Lecture | 4 | 5 | 4 | 10 | 25 | 48 | 3.0 |
| Programme | 6 | 3 | 4 | 7 | 28 | 48 | 3.0 |
| Total | 10 | 8 | 8 | 17 | 53 | 96 |  |

$x^2 < 9$   d.f. = 4   N.S. at 5% level

## TABLE VIII
### Question 7
#### Scores

|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total | Mean Score |
|---|---|---|---|---|---|---|---|---|---|---|
| Lecture | 3 | 4 | 1 | 5 | 4 | 10 | 6 | 15 | 48 | 5.0 |
| Programme | 7 | 3 | 1 | 2 | 4 | 6 | 10 | 15 | 48 | 4.8 |
| Total | 10 | 7 | 2 | 7 | 8 | 16 | 16 | 30 | 96 |  |

$x^2 < 12$   d.f. = 6   N.S. at 5% level

## TABLE IX
### Question 8
#### Scores

|  | 0 | 1 | 2 | 3 | Total | Mean Score |
|---|---|---|---|---|---|---|
| Lecture | 13 | 4 | 3 | 28 | 48 | 2.0 |
| Programme | 10 | 4 | 7 | 27 | 48 | 2.1 |
| Total | 23 | 8 | 10 | 55 | 96 |  |

$x^2 < 7$   d.f. = 3   N.S. at 5% level

TABLE X

Question 9

Scores

|  | 0 | 1 | 2 | 3 | Total | Mean Score |
|---|---|---|---|---|---|---|
| Lecture | 11 | 5 | 8 | 24 | 48 | 1.9 |
| Programme | 15 | 4 | 3 | 26 | 48 | 1.8 |
| Total | 26 | 9 | 11 | 50 | 96 |  |

$$x^2 < 7 \quad \text{d.f.} = 3 \quad \text{N.S. at 5\% level}$$

TABLE XI

Question 10

Scores

|  | 0 | 1 | 2 | 3 | 4 | Total | Mean Score |
|---|---|---|---|---|---|---|---|
| Lecture | 12 | 29 | 1 | 2 | 4 | 48 | 1.3 |
| Programme | 5 | 21 | 2 | 6 | 14 | 48 | 2.1 |
| Total | 17 | 50 | 3 | 8 | 18 | 96 |  |

$$x^2 = 11.76 \quad \text{d.f.} = 4 \quad \text{Sig. at 2\%}$$

As can clearly be seen from the foregoing tables only one question in the post-test detected any significant difference between the score distribution in the two treatment groups.

Table XI shows very clearly that in question 10 more "lecture" students had low scores than "programme" students; and fewer "lecture" students had high scores than "programme" students. If those students scoring 2 points are omitted, pooling gives the following distribution:-

|  | Low | High |
|---|---|---|
| Lecture | 41 | 6 |
| Programme | 26 | 20 |

As an alternative analysis a 't' test was carried out on the question 10 scores with the following results:-

| | Lecture | | | | Programme | | |
|---|---|---|---|---|---|---|---|
| $x$ | $f$ | $fx$ | $fx^2$ | $y$ | $f$ | $fy$ | $fy^2$ |
| 4 | 4 | 16 | 64 | 4 | 14 | 56 | 224 |
| 3 | 2 | 6 | 18 | 3 | 6 | 18 | 54 |
| 2 | 1 | 2 | 4 | 2 | 2 | 4 | 8 |
| 1 | 29 | 29 | 29 | 1 | 21 | 21 | 21 |
| 0 | 12 | 0 | 0 | 0 | 5 | 0 | 0 |
| | 48 | 53 | 115 | | 48 | 99 | 307 |
| | | $\Sigma fx$ | $\Sigma fx^2$ | | | $\Sigma fy$ | $\Sigma fy^2$ |

$$\bar{x} = \frac{53}{48} = 1.3 \qquad\qquad \bar{y} = \frac{99}{48} = 2.1$$

$$\Sigma' x^2 = 115 - \frac{53^2}{48} \qquad\qquad \Sigma' y^2 = 307 - \frac{99^2}{48}$$

$$= 56.479 \qquad\qquad\qquad = 102.813$$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\Sigma' x^2 + \Sigma' y^2}{48 + 48 - 2}\left\{\dfrac{2}{48}\right\}}}$$

$$= 3.607 \quad d.f. = 94 \text{ (Significant at 1\% level)}$$

In order to determine whether this question (Q.10) measured different outcomes from the others, each question was analysed on the basis put forward by Bloom et al (35). Within the cognitive domain there is a major distinction between objectives. There are those objectives (and consequently examination questions) which are based on the recall of information, which the students have learned, and those objectives (and consequently examination questions) which are based on the

intellectual skill of organising material either recalled or supplied. Those objectives which involve recall only are classified by Bloom as "knowledge"; those which involve intellectual skills are divided into five principal classes: "Comprehension"; "application"; "analysis"; "synthesis"; and "evaluation". This arrangement of behavioural objectives is regarded by Bloom as forming an hierarchy from simple to complex and not necessarily from easy to difficult. For example, if a learning situation were provided at the "knowledge" level and a question set which involved the intellectual skill "analysis of information", the student would probably find such a question difficult. If, however, a learning situation were provided at the "analysis" level and the students were asked a question which involved only "recall of information" then many would find this difficult.

The work of Cropp, Stoke, and Bashaw (36) has suggested that there is good reason for accepting the hierarchy although "evaluation" is seen by them as less complex than "synthesis" when the objectives are applied to science subjects. A recent publication by the Scottish Education Department (37) supports this view to some extent. It is pointed out in this document that, for science subjects it is difficult to separate "analysis", "synthesis", and "evaluation" the three being considered together as 'highest abilities'. For the purposes of analysing teaching situations or examination questions

on the basis of behavioural objectives, the objectives
have been coded thus:

A - knowledge
B - comprehension
C - application
D - highest abilities.

The number of marks awarded for each ability, in
each question, in the post-test was estimated from the
plan prepared for scoring the post-test.    The results
can be seen in Table XII.

TABLE XII

Taxonomic Structure of the
Post-test

|  | Proportion of marks awarded per objective | | | |
|  | A | B | C | D |
| Question 2 |  | 3 |  |  |
| 3 |  | 1 |  |  |
| 4 |  | 1 | 4 |  |
| 5 |  | 1 | 3 |  |
| 6 |  | 1 | 3 |  |
| 7 | 1 | 3 | 3 |  |
| 8 |  | 1 | 2 |  |
| 9 |  |  | 3 |  |
| 10 |  |  |  | 4 |
| Total | 1<br>3% | 11<br>32% | 18<br>53% | 4<br>12% |

The interesting feature of Table XII is the fact
that question 10 <u>does</u> appear to be different from the
others.    It is clear from the chi-square data that the
group taught by the programme did significantly better
on this question, than the group taking the lecture.    It
might very well be the case that the learning situations
provided by the programme are at the "highest abilities"

level whereas those provided by the lecture are not.
If this is so then questions testing those abilities will
be answered well by those students who followed the
programme.    The problem of relating teaching objectives
to testing objectives is more fully dealt with in a
subsequent chapter.

## CHAPTER 3

'"Contrariwise," continued Tweedledee, "if
it was so, it might be; and if it were
so it would be; but as it isn't, it ain't.
That's logic."'

<div align="right">Lewis Carroll.</div>

### Research on teaching methods

The primary aim of any investigation into teaching
methods should be to provide useful solutions to problems which
exist in the real world of education. Such research would
concern itself with those conditions, materials and
techniques which best facilitate students' achievement of
those objectives which are considered by teachers to be
desirable course outcomes.

The results of such researches should be presented
in such a way that they provide useful guidance for the
educationist, whether curriculum developer or teacher, in
ensuring that course objectives are achieved.

If this thesis is accepted there are three
principal factors to be considered:

(i) selection of the problems to be investigated,

(ii) the defining of variables involved, and

(iii) the control of experimental conditions.

(i)    When considering the problems to be investigated
it is important to discover whether method X is more effective
than method Y.   Cronbach (18) has argued that research on
teaching methods should endeavour to provide, not proof
that one method is more effective than another, but limited
generalisations concerning the effectiveness of particular
procedures or techniques.   In the present context, this
means that, instead of trying to establish that programmed

learning is more or less effective than lecturing, the problem to be investigated should concern the effects of this method on student intellectual behaviour for different subject areas or its success in promoting achievement of different kinds of objectives.

Cronbach argues against the search for widely applicable generalisations on the grounds that they do not exist. It is, he says simplistic to seek answers which could be generalised in a global way.

It is easy to see the force of such an argument. The results of research carried out on Heriot-Watt University students can be of little interest to those who teach chemistry say, in the U.S.A., since course content, student attitudes, academic standards, and so on may be vastly different in that country. Cronbach does not say, however, what he means by limited generalisation. Is it generalisation to other courses within the same institution or, as has been argued in Chapter 2, generalisation to other similar institutions ? Assuming that it is the second of these two possibilities then the scope of this present investigation meets the first of the three conditions referred to earlier, namely that the problem is correctly defined.

(ii) Again, as argued in Chapter 2 the independent and dependent variables have to be clearly defined if only to enable other investigators to repeat the study. The principal aim of teaching regardless of the method or technique employed, is to change student intellectual

behaviour. The stating of objectives is therefore one way of defining dependent variables. While this has not been explicitly done hitherto in this particular research the desired behavioural objectives are implicit in the questions set in the various tests. In order to produce some kind of response, in terms of student intellectual behaviour, the independant variable (i.e. the teaching method) is altered and observations made on the dependent variable (i.e. student intellectual behaviour).

(iii) The third factor, experimental control, has always been something of a dilemma for the educational researcher, although attempts have been made, particularly by Guba (38) to rationalise it. Guba's contribution has been effectually to distinguish between evaluation or development studies and research studies. An evaluator or developer, according to Guba, "needs to be able to refine and adjust his solutions continuously" and this of course could well upset the control which a researcher would require. In evaluation or development work conditions may be changed as the work proceeds whereas these conditions normally remain static in pure research. Thus it is possible that investigations into teaching methods may have acceptability criteria differing from those employed in other educational areas.

It does not necessarily follow, however, that to accept Cronbach's view is also to accept Guba's. It is possible to imagine a situation in which the criteria for

traditional research might well be violated to the
extent that several conflicting explanations of a final
result are possible, clearly casting doubt on the
validity of the study.    This kind of situation must
surely be avoided especially, if, as in Cronbach's view,
the results have to be practically useful.

In the physical sciences it is possible to define
research more precisely than in the behavioural sciences.
In addition the problem is all the more acute since the
word 'education' has several distinct meanings.    Two
definitions are given although it must be admitted that
neither is totally satisfactory.

> "Educational research may be defined broadly as
> any systematic striving for understanding activated
> by a need or sensed difficulty directed towards
> some complex educational problem of more than
> immediate personal concern and stated in problematic
> form." (39)

> "Thus we can say research (educational) should
> always denote careful, critical and exhaustive
> investigation to discover new facts which will
> test an hypothesis, revise accepted conclusions
> or contribute positive values to society in
> general." (40)

The first of these definitions makes the point that
research should be systematic and that it should arise as
the result of carefully considered problems.    The second
definition is in agreement with the first but includes an
extra factor;    it suggests that value judgments are
important, a point which has already been discussed.

## Interpreting the results

How are the results of an educational research
project to be interpreted ?    Great care has to be exercised

in the interpretation of data even in the most carefully designed experiments. It is possible, for example, for a teacher to become more favourably disposed towards a particular method simply because it is under investigation (the Hawthorne effect) and this change in the teacher's attitude could well influence the outcome of the experiment. The experimental and control groups may not, in fact, be strictly comparable in spite of efforts to make them so. The tests employed in the experiment may not be totally reliable or valid. When comparing experimental and control Groups, (even in a situation where comparability can be virtually guaranteed) the major difficulty is finding tests equally appropriate to both teaching methods.

These interpretative difficulties are, in essence, technical, but there are in addition logical difficulties such as the post hoc ergo propter hoc fallacy. If it can be shown, for example, that there is a strong positive correlation between marks scored in an examination and the amount of sleep a student gets, does it follow that lack of sleep has an adverse effect on the mind ? Or, if it can be shown that children who travel long distances to school do less well academically than those who live close to the school does it follow that it is a mistake to close village schools ?

Such difficulties as these, arise when a correlation coefficient or a difference between two means is shown to have statistical significance.

Conversely, if a statistical test shows that a correlation or mean difference is not significant it does not always follow that no association exists, or that the two groups are not in fact behaving in a like manner.

Failure of a result to attain statistical significance could arise because the sample was too small or because the post-test used was not sufficiently valid or reliable to detect any difference which exists. Indeed it could well be the case that the post-test was measuring entirely the wrong attributes.

In both the experiments described in Chapter 2 no statistically significant mean difference was detected between Group A taught by one method and Group B taught by another. This was still the case even when variables which were difficult to control experimentally, were controlled statistically using analysis of covariance.

Apart from the obvious weakness of sample size in Experiment 1 there is always the possibility that the tests used in research of this kind are of questionable validity. Suppose, for example, that the post-test in Experiment 2 did not measure abilities within the subject area dealt with in the teaching, but measured instead some other attribute such as mathematical ability. If the experiment had been well designed and the experimental and control groups were indeed comparable, a non-significant value of F would have been obtained, due to an inappropriate or invalid test being used. Most research studies on the effectiveness of teaching methods do in fact measure outcomes on the results of a single post-test, following the 'treatment'

and in the majority of cases too little attention is paid to the validity of the post-test.

It is impossible to say whether the factors which influence learning from one technique, such as programmed learning are the same as those which influence learning from formal teaching. There may very well be, for example, social factors operating (although it must be admitted that this is less likely to be the case in a university than in a school) which determine outcomes. This kind of difficulty was underlined in the Plowden Report (41) where a relationship between pupil achievement and parental interest was reported. The criterion of success used in this work was reading ability assessed by a reading test. While reading ability is an important outcome of primary school education it is not a valid test of total achievement. Had a test of numerical ability been used instead it is probable that no relationship with parental interest would have been claimed. The conclusion then is questionable as a result of using an invalid test of achievement.

For the reasons outlined above, it is a serious misuse of research findings to conclude that Method A is a more effective way of teaching than Method B on the basis of one or even two post-tests the validity of which is doubtful.

A much more dangerous misuse of results is the extrapolation of findings beyond the population on which the study was carried out. In this present study it has been claimed that given a valid post-test the results may be generalised beyond the confines of Heriot-Watt University to first year courses at other Scottish Universities.

What is not justified is extrapolation to non-university educational situations. The results of both Experiment 1 and 2 seem to point in the same direction. Such a result increases the confidence with which a generalisation may be made, and certainly a generalisation of some description is probably justified. Generalisation, even if it is pure conjecture, is justified only if it results in further research.

Limited enquiry such as described in this thesis can seldom, if ever, produce definite answers to definite questions. What it should do is stimulate thought, refine the questions, and refine the experimental design. The research outcome, then, is the stimulus to further enquiry rather than a final answer to a question.

Having argued that the results of a research project, particularly a limited one such as this, should be used as a stimulus to re-thinking leading to further study, the problem of how this re-thinking is to be organised must now be considered. This is probably best done under a number of heads. These are:-

1. The post-test in comparative experiments.
2. The Lecture and the learning process.
3. Programmed instruction and the learning process.
4. The evaluation of programmed learning research.
5. The problem of comparisons.
6. Unsolved problems.

1. The post-test in comparative experiments

Cronbach, as indicated earlier, has pointed out that research into teaching methods should not involve comparative studies but should strive to examine techniques and estimate their effectiveness. While, no doubt, a

case may be made for this point of view, comparative
studies are <u>not</u> invalid, provided various criteria are
met.

In the physical sciences success in comparing changes
in one situation with changes in another depend to a large
extent on the sensitivity of the measuring instrument.
This is also the case in the behavioural sciences.  The
lack of measurable difference between the programmed
learning technique and the lecturing technique might have
been due to a lack of sensitivity in the post-test, to
 measure any differences which did exist.   It must be
admitted, of course, that differences which are
statistically significant may not be of practical importance.
Given large enough samples, the most trivial difference
becomes statistically significant.

The use of routine class tests, either made up by
the lecturer (the case in this investigation) or published
by some organisation such as the American Chemical Society
is suspect.   Tests of this type are usually designed
simply to differentiate among individual learners and are
invariably insensitive to the type of learner growth which
might take place in one group but not in another as a
result of different teaching methods being used.

The crucial problem, then, is the construction of a
post-test which is sensitive enough to detect small
differences in effectiveness between two methods.   If, as
stated earlier, the principal aim of teaching, regardless
of method, is to bring about changes in the students'
knowledge, attitudes and skills, then it is the behavioural

objectives prescribed prior to the teaching process that
the post-test has to measure.

In other words, before any teaching takes place the
behavioural objectives involved must be listed and these
must be the same for both methods, which necessitates
that, in an experiment comparing lecturing with programmed
learning the lecturer involved also writes the programme.
The post-test, when constructed, must measure these
pre-specified objectives, and no others.   To summarise
then there are four stages involved in constructing a
highly sensitive test of this type.

First, the teacher or lecturer decides which objective
or objectives are appropriate and makes a clear and explicit
statement describing the objectives in behavioural terms.

Second, the lecture is planned to cover the same content
material as the programme with the same number of illustrative
examples and in the same detail as the programme.

Third, the two groups are exposed to the two methods.
If the subject material is not totally unfamiliar to the
students a pre-test is necessary to exclude the possibility
that the intellectual behaviour described by the objectives
is not already within the student's accomplishments.
If the objectives are based on an unfamiliar topic then
the pre-test can be dispensed with.

Fourth, the post-test is administered, measuring only
the attainment of those behavioural objectives specified
at the start of the experiment.

The foregoing argument is concerned only with those

behavioural objectives described within the cognitive domain. An important, if perhaps qualitative estimate of teaching effectiveness could be obtained by including in the post-test items which measure objectives in the affective domain. To ask the question, "Which of the two teaching methods, lecturing or programmed learning do you prefer?" is probably not enough. A more searching question would be "Did you find the lecture/programme on infrared spectroscopy interesting?" or "Did you enjoy the lecture/programme on ionic equilibria?". It is important to make some kind of assessment of this sort since a teaching procedure could well achieve the cognitive domain objectives at the expense of the affective domain. In other words the student is taught to learn chemistry but also to hate it.

Unless the precautions listed in these four points are adhered to the post-test will measure objectives not involved in the teaching processes under investigation, and will be reduced in sensitivity. Reduction of sensitivity in this way could very well explain the non-significance of results obtained in both experiments described in Chapter 2.

This theory is supported by the analysis of question 10 (see Appendix V) which does detect a difference between the two methods. The Chi-square calculations for this question (see Table XI page 57) gives a value of 11.76 which is significant at the 2% level clearly indicating that those students who studied the programme scored significantly better on this question. Again, as shown

in Table XII (page 60) this is the only question in
the test which examines abilities of a high order of
complexity.   The other nine questions are concerned
with less complex objectives (mainly Groups B and C)
and none has a significant Chi-square.   It seems
reasonable to assume therefore that the programme provided
learning situations at the complex end of Bloom's
hierarchy whereas the lecture did not.

Clearly, if research into the effectiveness of
different methods of teaching is to produce useful results
more attention will have to be paid to the behavioural
objectives involved, and objective testing will have to
be used for the evaluation of outcomes.   Only in this
way will the post-test be valid and be sensitive enough
to detect the small differences expected between methods.

It is always possible that, these precautions stated
above notwithstanding, a false result could be obtained,
either because the post-test was not as valid as it was
thought to be or because the lecturer's performance fell
short of the level he usually attained.   These two
difficulties could be overcome by allowing the lecturer
to give a series of lectures (compared with a series of
elements from the one programme) and have more than one
post-test covering the series.   Any atypical performance
or assessment is then allowed for, and a more representative
estimate of the effectiveness of the techniques obtained.

2.  The lecture and the learning process

A student's preference for any particular teaching

technique for another depends, to some extent, on whether
he thinks he learns more from the technique of his choice.
The effectiveness then or otherwise, of lecturing depends on
what the student wants from the lecture.  The majority
of students consider it of paramount importance that the
lecturer should know his subject and be able to discuss
it clearly and unambiguously.  Lecturing style is also
considered important, and the lecture should be well con-
structed and relate properly to other lectures in the
same course.  Smithers (42) has pointed out that what
is required of a lecturer depends to some extent on the
personality of the student.  He found, for example, that
extraverts favoured lectures which were a "performance"
more than did introverts.  According to the extraverts
the ideal lecturer was also entertaining and self-confident.
Neurotic differences within student groups seem principally
to be associated with lecture content rather than lecture
style.  Those students who are high on neuroticism, and
therefore anxiety, prefer lectures which are highly specific
from which the student can make full and complete notes for
examination purposes.

Furneaux (43) has produced evidence to support this
view and has shown that neurotic-introverts are generally
more successful in examinations than other students.

There is, then, ample evidence which suggests, that
the success or otherwise of lecturing as a teaching method
depends to some extent on the personality of the student.
The student audience does not consist of two hundred
replicas of the same individual but of groups of types of

individual each requiring different attributes from the
lecturer.

Might there be a similar relationship between style
(or content) and student personality for all forms of
teaching, including programmed learning ?

3. Programmed instruction and the learning process

It has always been emphasised by the protagonists
of programmed instruction that the important feature
of this teaching technique is the programme itself.   There
seems little doubt that programme writers give considerable
thought to the aims of the programme but very little to
objectives.   Apart from these theoretical educational
considerations there are other more practical issues which
require thought.   What age should the students be ?
Is the programme for men or for women ?   Is it written for
bright, dull or average students ?   Alas there seems no
evidence that student personality has ever been considered.

A number of educational advantages are, however,
claimed for this type of instruction.

First, when a student works through a programme he
is actively taking part in a process whereas, if he is
listening to a lecture he is, at best, active in a spasmodic
way only.

Second, programmed instruction is fundamentally
individual teaching.   With Crowderian programmes, as
explained in Chapter 1 the information sequence presented
to the student depends on the students, on responses and
so, to this extent, the programme is adapting to individual

needs.   For both Skinnerian and Crowderian programmes
the student works at his own pace.

Third, answers are immediately indicated as correct
or wrong which is (according to Skinner) important for
learning.

Fourth, by its very nature, more planning goes into
the preparation of a programme than goes into the
preparation of a lecture.

Fifth, since programmes (like objective tests) have
to go through a validation process involving much rewriting
the end-product will be a programme which works (at least
for a proportion of students) whereas there is never any
guarantee that a lecture will achieve its aims.   Most
teachers take the view that if students fail, it is their fault
and not the lecture's.   This cannot happen with programmed
learning.

When programmed learning is used in a purely educational
way (i.e. not as part of an experiment) it is important
to stress that in such a situation it is the student who
controls the rate of work and the way in which he reacts
to the responses he makes.   The programme writer's desire
to motivate the student through a low error rate while
perhaps desirable, is certainly not enough;   nor is it
enough to prevent him from looking at the correct response
before he answers the question.   It is surely good
educational practice to allow the student to re-read a
passage if he so wishes, or to look at an earlier section
previously completed;   or to inspect a section not yet
arrived at.   Such 'freedom' is allowed by the programmed

text but not by the teaching machine.

There can be little doubt that, while factual subjects like the physical sciences lend themselves to programmed teaching much more readily than subjects such as philosophy, or music appreciation, (where value judgments and personal taste are important) the science student dislikes the programmed approach.

While it was outwith the scope of this investigation to examine the reasons for this dislike, the hypothesis that programmed learning is isolated from the materials and equipment of science (and consequently tedious and dull, compared to a lecture, which might be illustrated with demonstrations of this reaction or that property) seems a reasonable explanation.

If chemistry programmes were written to be used in the laboratory (rather than in the study or the library) in conjunction with apparatus, chemicals, specimens, laboratory reference texts and so on, the students' opinion might be totally different from that usually reported.

The emphasis on planning, testing and revision of programmes is, as already pointed out, a highly desirable operational scheme, ensuring that the final programme is valid. As with the case of objective testing the various operational criteria laid down are perhaps educationally suspect. For example, is the belief in a five per cent error rate for linear programmes based on experimental evidence or is it simply Skinner's opinion ? How important is it to require overt responses ? How

important is it to require immediate knowledge of results ?

Put in another way, the systems, methods and techniques used by programmes are too exact, too clinical, and too educationally unsophisticated for an essentially complex learning process.   It is not surprising therefore that many students (even clinical, exact scientists) find programmed texts dull, uninteresting and devoid of imagination.

The possible exception to the generalisation might be the neurotic intravert or extreme converger .   As Goldstein and Gotkin (44) have pointed out:

> "The procedure of developing testing and doing
> basic research with programmed materials is
> unhappily devoid of teacher involvement except
> in the instances where teacher and programmer
> are one and the same."

4.   The evaluation of programmed learning research

The writing of a programme is a piece of work to which the programmer has to devote his full-time attention. Consequently if a commercially available programme can be used in research then it should be, since the skills involved in writing and validating a programme are not possessed by the majority of lecturers.   This difficulty can be overcome by training lecturers to write their own programmes.   The two programmes used in the experiments described in Chapter 2 are both commercially available but were written in part by Dr M. M. Campbell, of Heriot-Watt University Chemistry Department who acted as consultant for all aspects of the use of programmes.   The lecturers involved (Dr I. Soutar - Experiment 1 and Mr R. B. Snadden -

Experiment 2) were selected for their known enthusiasm
for teaching and their interest in chemical education
research.

The main difficulty, concerning individual lecturers,
involved in experimental situations is one of representative-
ness, not only of the whole population of lecturers but
also of the lecturers' own behaviour pattern.  Can it
be ensured that Dr X is representative of chemistry
lecturers as a whole ?  If he is outstandingly good (as
some surely are) or abysmally poor (as many are) does
this not unfairly bias the outcome of any experiment
involving Dr X ?  Similarly, is the programme produced
by Dr Y representative of all available programmes ?   If
it is much better or much worse than most does this not
bias the outcome ?  A great deal of effort has been
made in this study, and in most others to ensure
representativeness among the students but not lecturers
on the programmes.

5.  The problem of comparisons

In most practical situations it is unlikely that
any real comparison can be made between a pure programmed
treatment and a pure non-programmed treatment.  It would
be difficult in normal circumstances to isolate teaching
technique X from technique Z and the most which can
probably be achieved is comparing $(X + Y)$ with $(Y + Z)$
where Y is another factor influencing teaching.

As explained in Chapter 2 good experimental design
ensures that such variables as difficulty of subject
material, time spent in teaching, (or other independent

variable associated within teaching) should be equalised
between the two techniques.   It is difficult nevertheless
to ensure that the two situations compared are different
only in the _way_ they are taught.   Experimental tests
of the effectiveness of one technique compared to another
must, for the moment, remain suspect.

6.  Unsolved problems

Clear-cut solutions to many of the problems associated
with self-instruction are unlikely to be provided for
reasons that have been dealt with earlier.   It is possible,
however, to indentify those areas where possibility of
solutions seem greatest.

While the system of communication in self-instruction
remains as some form of programmed text, success or failure
of this method of teaching probably depends on four
principal factors.   These are:-

        (a)  The learner's motivation and possibly
             his intelligence,
        (b)  The type of teaching situation in which
             the programme is used,
        (c)  The programme quality,
        (d)  Future research.

(a)  Motivation and intelligence

Modern educational thinking does not consider
intelligence as a fixed mental capacity, nor does it
consider that high ability is in short supply.   Further,
educational theory no longer subscribes to the view that
motivation to learn is located entirely within the
individual, but can be produced by external social factors.

The philosphy which advocates the use of programmed
learning as a teaching method is in agreement with both

these points of view.

As far as intelligence is concerned it is well known that bright students learn more than dull ones no matter what teaching method is employed, but when programmed learning is used as the teaching method, the not-so-bright do almost as well as the bright (at least at school level if not at university level).

The implication is that the logical and small step sequence of the linear programme counteracts the fear of failure and enables the majority to maintain enthusiasm and attention. It has been claimed by Leith (45) that a significant correlation is obtained between I.Q. and test scores when traditional methods of teaching are used but this is not the case when programmed learning is the teaching method.

> ".... in the classroom teaching situation the
> learners task is made more difficult because
> he must unscramble what the teacher says.
> This rearrangement and perceiving of what is
> relevant can be done by the intelligent but not
> by the less intelligent."

There still remains the problem of the bright student who is lazy or indifferent; how does he react to programmed learning ? Are the students who are indentified by Hudson as divergers, excelling in the open-ended situation, adversely affected by programmed learning ? Do the controlled thought processes employed during a programmed learning session militate against imagination or creativity ? Is it ever possible to be sure that even if the method is good for the many it is not positively bad for the few ?

(b) The teaching situation

It is clear that, for the present, the way in which

programmes are to be used must be left to the individual
lecturer or teacher to decide.   While programmed learning
was introduced originally to act as a substitute for formal
teaching it would be inadvisable at the present moment
totally to replace conventional teaching.   Such replacement
may well become a real possibility in the future when prog-
rammes are more refined and techniques for testing their
outcomes more sensitive.

Until such time, when it can be shown unequivocally
that programmed learning is more effective than lecturing,
there are many ways in which the technique can be put to
good effect.   These include revision work, starting
points for discussion, and private home study assignments.
If the results of American research are accepted (46) a
higher attainment standard is achieved if students are
given the opportunity for self-instruction, at home,
subsequent to the lecture.

(c)  Programme quality

Provided that a linear programme has been properly
validated it can cater for a very wide range of abilities
allowing each individual to work at his own pace.   As
pointed out earlier there is evidence to suggest that both
bright and dull students achieve a similar standard of
achievement.   It is difficult to believe, as some
educationists would have it, that programmed learning
is in the above sense, an egalitarian influence.   To do
so is to accept Mill's dictum that "The generality of
children are organised so nearly alike that they may by
proper management, be made pretty nearly equally wise and

virtuous.'

In some ways programmes have student weakness built
into them.   The writer, obsessed with the belief that
the ideal error rate is something less than five per cent,
will of necessity, produce a programme, with which the less
able student can learn significantly.   It can never be
assumed that such a programme does not have an adverse
effect on brighter students.   It is probably a case
of all learning the same amount but some learning a lot less
than they might have done had some other teaching method
been used.   This criticism of programmed learning is
certainly justified in a primary school or in a mixed
ability class in a secondary school, but it is doubtful if
it applies in a university situation where selection for
places makes the ability range fairly narrow.

It remains clear that if programmed learning is to
be effective at all, the programme has to be constructed,
validated, and used with the particular needs of particular
students in mind.   As Gotkin has pointed out,

"To pass through a single instructional path at
ones own rate cannot be equated with the tutorial
situation. To argue otherwise is to offer a naive
notion of individualised instruction.   By enabling
students to proceed at their own pace, programmed
instruction <u>does</u> break the traditional lockstep of
classroom procedure. In breaking the lockstep it
makes an enormous stride forward in individualising
instruction.   But that is only one dimension of
individualisation (47)".

(d)   The future

While much research has now been carried out on the
various problems associated with programmed learning, such
as overt versus covert responses, or constructed response

versus multiple choice, or simply programming versus
traditional teaching, very little data have been produced
which confirm one view or the other.    As pointed out in
Chapter 2, if an hypothesis requires to be tested by
an experiment then variables have to be identified and
held constant within the experiment.    In any teaching
situation the number of variables which might affect the
outcome is very large indeed.    Fry (48) illustrates
this point by listing some twenty-six classes of variable.

With such a large number of interacting variables it
is almost impossible to control an experimental situation
so that the results will not be contradictory or
inconclusive.    More carefully designed research is required
if this problem is ever to be overcome.    To quote Holland (49)

> "The need is for constructive research which clarifies,
> defines and quantifies the parameters of programs.
> The frequent use of "versus" in the titles of studies
> is symptomatic of an inability to measure variables
> even approximately.    How sequenced is a programme
> comparing "sequenced vs non-sequenced" arrangement
> of items ?    When one can, at least roughly, identify
> in a program the magnitude of each of several
> dimensions, the technology of programming will be
> much further along, and "versus" will be a thing
> of the past."

It is often argued that any subject which can be taught
can be programmed, but it does not follow that programming
is the best way of teaching it.    From the evidence
reported in Chapter 2 and reviewed in Chapter 1 the case
for or against programmed learning must remain "not proven."

## ACKNOWLEDGMENTS

The author wishes to express his most sincere thanks to Dr A.E.G. Pilliner for his meticulous supervision of this research work, for his many valuable suggestions and constant encouragement.

Best thanks are also due to Professor B.G. Gowenlock, for allowing the experimental work to take place within his department, to Drs. Campbell, Snadden and Soutar for their co-operation and interest and to the students for the cheerful way in which they submitted themsleves to lecture, programme and test.

Sincere thanks are due to Dr D. M. McIntosh for his advice and help and to the Governors of Moray House College of Education for financial support, without which there could have been no study.

## REFERENCES

(1)   Pressey, S.L.
      A Third and Fourth Contribution Towards the Coming
      Revolution in Education
      School and Society 1932.

(2)   Skinner, B.F.
      The Science of Learning and the Art of Teaching
      Harvard Educational Review 1954.

(3)   Crowder, N.
      Automatic Tutoring by Means of Intrinsic Programming
      Automatic Teaching:  The State of the Art.
      (E. Galanter - Ed)  Wiley Hill 1959.

(4)   Cavanagh, P., and Jones, C., (Eds)
      Programmes in Print
      Association for programmed learning.
      (London) 1966.

(5)   Schonell, F.J., et al
      University Teaching in Queensland:  A Report of
      Conferences for Demonstrators and Lecturers
      University of Queensland, Brisbane 1961.

(6)   University Grants Committee
      Report of the Committee on University Teaching
      Methods (Hale Report)
      H.M.S.O. 1964.

(7)   Joyce, C.R.B., and Weatherall, M.
      Controlled Experiment in Teaching
      The Lancet, $\underline{2}$, 402, 1957.

(8)   McLeish, J.
      The Lecture Method
      Cambridge Monographs on Teaching Methods, No.1
      Cambridge Institute of Education 1968.

(9)   Elton, L.R.B., Hills, P.J., and O'Connell, S.
      Teaching and Learning Systems in a University Physics
      Course
      Phys. Educ. $\underline{6}$, 95, 1971.

(10)  Elton, L.R.B., Bond, D.J., Nuttall, J., Stace, B.C.
      Teach yourself pardigm:  the Keller Plan
      Chem. in Brit., $\underline{9}$, (4), 164, 1973.

(11)  Keller, F.S.

      J. App. Behav. Anal. $\underline{1}$ 79 1968.

(12)  Green, B.A.,
      Physics teaching by the Keller plan at M.I.T.
      Am. J. Phys., $\underline{39}$, (7), 764, 1971.

(13)  Gunstone, F.D., and Moyes, R.B.,
      <u>Programmed Learning in Chemistry I - First Principles</u>
      Educ. in Chem., <u>1</u>, (4), 189, 1964.

(14)  Hoare, D.E., and Inglis, G.R.
      <u>Programmed Learning in Chemistry, II - An Experiment</u>
      Educ. in Chem., <u>2</u>, (1), 32, 1965.

(15)  Young, J.A.,
      <u>Programed Instruction</u>
      J. Chem. Ed., <u>40</u>, (1), 11, 1963.

(16)  Hogg, D.R.
      <u>The Use of Programmed Learning with Second Year
      University Students</u>

      In <u>Aspects of Educational Technology</u>,
      D. Unwin and J. Leedham (Eds) p.315
      London: Methuen, 1967.

(17)  Hogg, D.R.
      <u>Student Attitudes to Programmed Learning</u>
      Ed. in Chem., <u>10</u>, (1), 7, 1973.

(18)  Cronbach, L.J.
      "<u>The Logic of Experiments on Discovery</u>,"
      <u>Learning by Discovery: A Critical Appraisal</u>
      (Shulman, L., and Keisler E., Eds)
      Rand McNally Chicago 1966 pp 77-96.

(19)  Wallis, D. and Wicks, R.P.
      <u>The Royal Navy Study</u>
      Occup. Psychol., <u>37</u>, (1), 50, 1963.

(20)  Knight, M.A.G.,
      <u>The Royal Air Force Study</u>
      Occup. Psychol., <u>37</u>, (1), 68, 1968.

(21)  Cavanagh, P., Thornton, C., and Morgan, R.T.G.
      <u>The British European Airways Study</u>
      Occup. Psychol. <u>37</u>, (1), 76, 1963.

(22)  Smith, N.H.
      <u>The Teaching of Elementary Statistics by the
      Conventional Classroom Method Versus the Method of
      Programmed Instruction</u>
      J. Educ. Res., <u>55</u>, (9), 417, 1962.

(23)  Green, E.J., Weiss, R.J., and Nice, P.O.
      <u>The Experimental Use of a Programmed Text in a
      Medical School Course</u>
      J. Med. Educ., <u>37</u>, 767, 1962.

(24)  Elton, L.R.B., Hills, P.J., and O'Connell, S.
      <u>Teaching and Learning Systems in a University Physics Course</u>
      Phys. Educ., <u>6</u>, (2), 98, 1971.

(25) Hughes, J.L., and McNamara, W.J.
Comparative Study of Programmed and Conventional
Instruction in Industry
App. Psychol., Aug. 1961 p.225.

(26) Ferster, C.B., and Sapon S.
An Application of Recent Developments in Psychology
to the Teaching of German
In Teaching Machines and Programmed Learning
(Lumsdaine, A.A., and Glaser R., Eds) Washington, D.C:
N.E.A., 1960, pp. 173-185.

(27) Mackaill, A.W.
The Nature and Measurement of Chemical Ability
Unpublished Doctoral Dissertation.

(28) Campbell, D.T. and Stanley, J.C.
Experimental and Quasi-experimental Designs for
Research in Teaching
In Gage, N.L. (Ed) Handbook for research in Teaching.
Rand McNally 1963 pp171-296.

(29) Lindquist, E.F.
Design and Analysis of Experiments In Psychology
and Education
Houghton Mifflin (Boston) 1956 p 54.

(30) Creswell, C.J., Runquist, O., Campbell, M.
Spectral Analysis of Organic Compounds
Burgess (1972).

(31) Johnston, A., Morrison, T., and Sharp, D.W.A.,
Topic Difficulties in Chemistry
Educ. in Chem. $\underline{8}$, (6), 212, 1971.

(32) Wechsler, D.
The Measurement and Appraisal of Adult Intelligence
Bailliere, Tindall and Cox 1958.

(33) Runquist, O., Creswell, C.J., and Head, J.T.
Chemical Principles a Programmed Text, parts I and II
Burgess Publishing Co., Minneapolis 1968.

(34) Lindquist, E.F.
Design and Analysis of Experiments in Psychology
and Education
Houghton Mifflin (Boston) 1956 p 317.

(35) Bloom, B.S. (Ed)
Taxonomy of Educational Objectives
Vol. I Longmans 1956.

(36) Cropp, R.P., Stoke, W.H., Bashaw, W.L.
The Construction and Validation of Tests of the
Cognitive Process as Described in the Taxonomy of
Objectives
Co-operative Research Project No.2117, Dept. of Educ.
Res. and Testing. Florida State University 1966.

(37) Scottish Education Department
Curriculum Paper No.7 - Science for General Education
H.M.S.O. 1969 p 27.

(38) Guba, E.
Significant Differences
Educ. Researcher, 20, 4, 1969.

(39) Harris, C.W. (Ed.)
The Encyclopaedia of Educational Research
3rd Edition 1960 p 1160
(MacMillan).

(40) McAshen, H.H.
Elements of Educational Research
McGraw-Hill (New York) 1963.

(41) Plowden Committee
Children and their Primary Schools
H.M.S.O., London 1967.

(42) Smithers, A.
What do Students Expect of Lectures ?
Univ. Quart., 24, 330, 1970.

(43) Furneaux, W.D.
The Psychologist and the University
Univ. Quart., 17, 33, 1962.

(44) Goldstein, L.S., and Gotkin, L.G.
Teaching Machines versus Programmed Textbooks
J. Prog. Instr., 1963.

(45) Leith, G.O.M.
Research on Programmed Learning
(A Handbook of Programmed Learning, Birmingham
University) p 83 1964.

(46) Gotkin, L.G., and Goldstein, L.S.
School Utilisation of Programmed Instruction:
Implimentation Studies
(C.P.I.) 1962.

(47) Gotkin, L.G.
Programmed Instruction in the Schools:
Individual Differences, the Teacher and Programming
Styles
(C.P.I.) 1962.

(48) Fry, E.B.,
Teaching Machines and Programmed Instructions
McGraw-Hill (New York) p 195, 1963.

(49) Holland, J.G.,
Research on Programming Variables
in Teaching machines and Programed Learning (Glaser, R.,
Ed.)
National Education Association of the United States
1965 p 66.

# APPENDICES

# A P P E N D I X   I

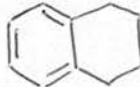## Spectroscopy  Tests

### Section A    Ultraviolet Spectroscopy

1. The quantised energy of a molecule is comprised of electronic vibrational and rotational energy contributions. Indicate the type of energy change which is <u>primarily</u> responsible for U.V. spectra.

2. Which types of transition (i.e. $\sigma - \sigma^*$ etc.) are possible in the following molecules.   State which of these will be the transition of longest wavelength in each case.

   a) $CH_3NH_2$    b) $C_2H_5SC_2H_5$    c) $CH_2=CHCH_3$    d) $CH_3C\overset{O}{\underset{H}{\diagdown}}$

3. Why are special precautions necessary for the examination of transitions whose wavelengths lie below 200 um ?

4. Draw an energy level diagram to show how conjugation in the system $>C=C-C=C<$ affects the energy of the $\pi-\pi^*$ transition relative to that in the isolated $>C=C<$ case. How will the wavelength of the transition be affected ?

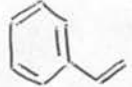5. Place the following compounds in increasing order of $\lambda_{max}$ for the $\pi-\pi^*$ transition?
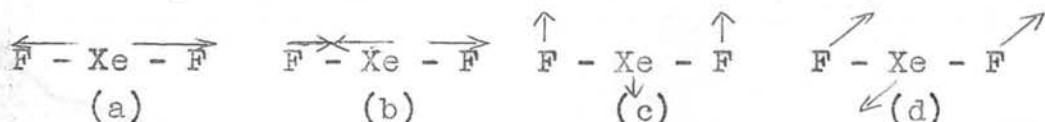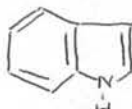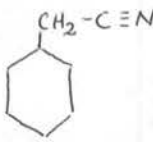
   (a)          (b)          (c)          (d)

6. Compound A shows a $\lambda_{max}$ at 240 nm in water and at 260 nm in cyclohexane.   What kind of transition is involved ? Show by means of an energy level diagram why the solvent affects the transition in this manner.

## Section B     Infrared Spectroscopy

1.  How many vibrational degrees of freedom are possessed by a
    molecule of Water ?

2.  The linear molecule xenon difluoride shows the vibrational
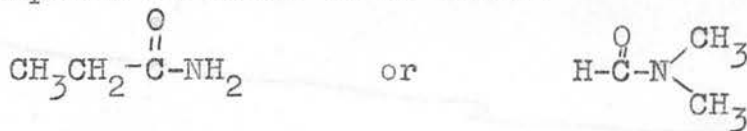    modes depicted below

    $\overleftarrow{F} - Xe \overrightarrow{- F}$     $F \overset{\times}{-} Xe - \overrightarrow{F}$     $F \overset{\uparrow}{-} Xe - \overset{\uparrow}{F}$     $F - Xe - F$
         (a)               (b)               (c)               (d)

    (I)  which of these will be infrared active ?

    (II) what is meant by the statement that modes (c) and (d)
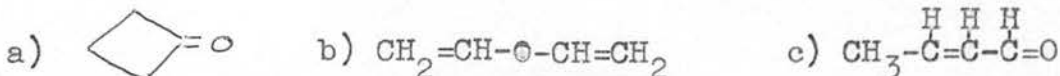         are degenerate ?

3.  A compound is known to be a)  [indole structure]  or b)  [cyclohexane with $CH_2-C\equiv N$ substituent]

    Use the spectrum shown on the attached sheet to distinguish
    between the two.

4.  A compound is known to be either

    $CH_3CH_2-\overset{\overset{O}{\|}}{C}-NH_2$          or          $H-\overset{\overset{O}{\|}}{C}-N\overset{CH_3}{\underset{CH_3}{<}}$

    Which region of the infrared spectrum of the compound could
    be used to decide which of the two possible structues is
    correct ?  (state reason)

5.  A compound of formula $C_4H_6O$ shows absorption in the region
    $3300 - 2900$ cm$^{-1}$.  Absorption in the region $1900 -1650$ cm$^{-1}$
    is evident.   Simple Simon has suggested the three structures
    below.

    a)  [cyclobutane]$=O$     b) $CH_2=CH-O-CH=CH_2$     c) $CH_3-\overset{H}{\underset{}{C}}=\overset{H}{\underset{}{C}}-\overset{H}{\underset{}{C}}=O$

    Only one of these structures affords a possible solution.
    Can you help Simon decide which is correct ?

INFRARED SPECTRUM. QUESTION 3.

a)

b)

WAVENUMBER (cm$^{-1}$)

4000    3500    3000    2500    2000
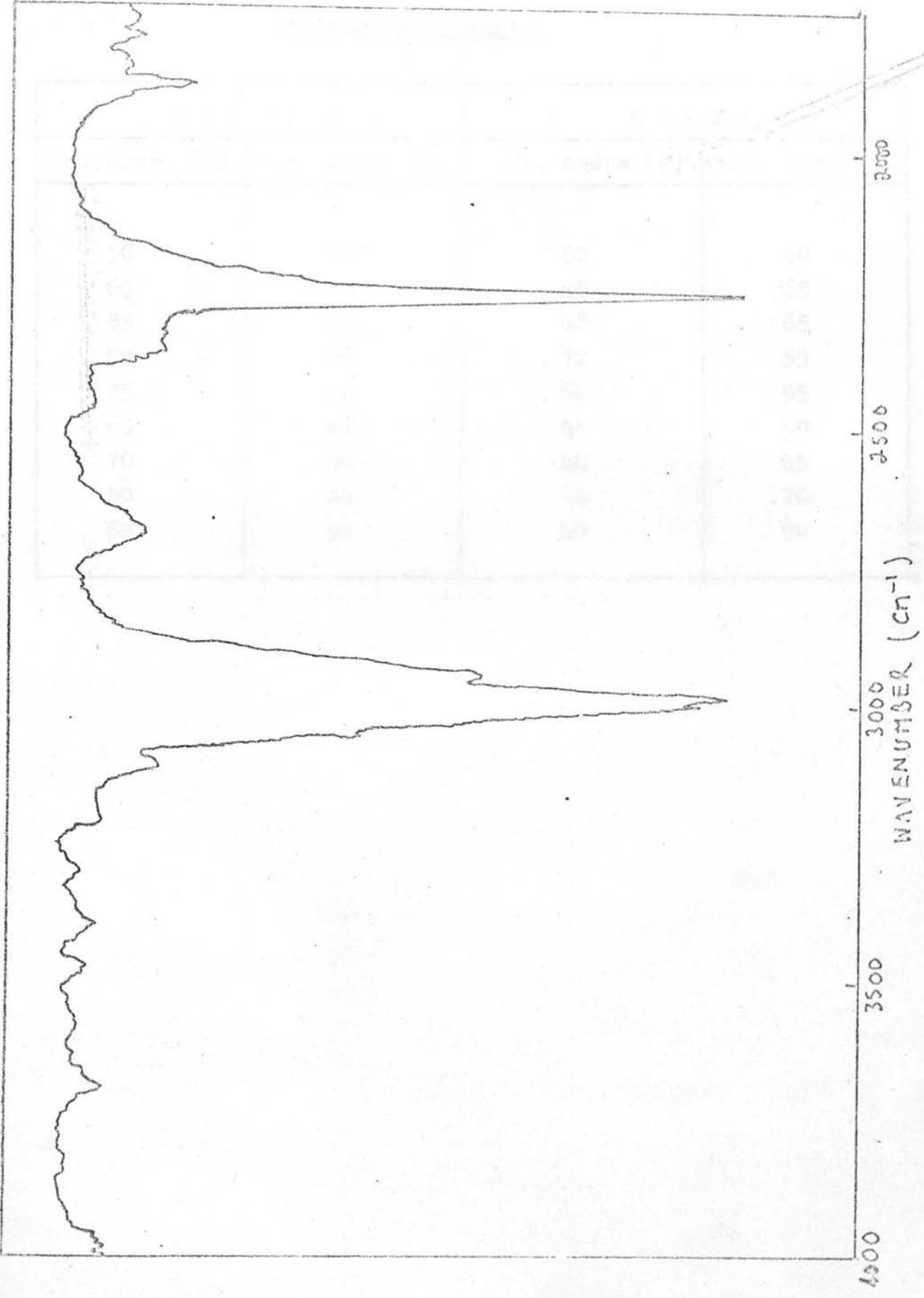
# APPENDIX  II

## Data for Experiment 1

| G R O U P  A | | G R O U P  B | |
| --- | --- | --- | --- |
| i.r. score (P) | u.v. score (L) | u.v. score (P) | i.r. score (L) |
| 50 | 64 | 52 | 60 |
| 65 | 60 | 40 | 65 |
| 85 | 52 | 48 | 85 |
| 90 | 56 | 72 | 55 |
| 75 | 48 | 64 | 95 |
| 65 | 44 | 44 | 60 |
| 70 | 40 | 44 | 45 |
| 50 | 24 | 64 | 70 |
| 69 | 92 | 40 | 90 |

## APPENDIX III

## Randomisation Programme

| Log Drive | Cart Spec | Cart Avail | Phy Drive |
|-----------|-----------|------------|-----------|
| 0000 | 0005 | 0005 | 0000 |
| V2 M11 | Actual 8K | Config 8K | |

```
// for
* Iocs(Card,1132 printer, disk, keyboard, typewriter)
* list source program
* one word integers
    dimension N (138), M (2)
    IX=3991
    Do 10 I=1,138
10  N (I)=I
    Do 20 I=1,69
    Do 30 J=1,2
40  call randu (IX,IY,Y)
    IX=IY
    NPT=Y*138.0+1.0
    IF(N(NPT).EQ.0)GOTO 40
    M(J)=NPT
    N(NPT)=0
30  continue
    write (3,50)M
50  format ('',2I6)
20  continue
    call exit
    End
    Features supported
    one word integers
    IOCS
    Core requirements for
    Common        0 Variables      150 Program      150
End of compilation
```

## APPENDIX IV

### The Pretest

1. (a) Fill in, on the blank Periodic Table supplied, an
   example of the undernoted type of elements. Use
   the symbol of the element you select with a
   superscript corresponding to the question number.
   The same element may be used more than once.

   (1) A gas ................. $H^1$

   (1) An alkali metal
   (2) A halogen
   (3) An active non metal
   (4) An active metal
   (5) An element with a basic oxide
   (6) An element with an acidic oxide
   (7) An element with an amphoteric oxide
   (8) An element which exhibits variable valency
   (9) A reducing agent
   (10) A member of the first short period
   (11) A member of the second short period
   (12) A liquid
   (13) A monatomic gas
   (14) An element which exhibits allotropy
   (15) An element which forms soluble salts
   (16) An element which forms a stable oxide.

   (b) Very briefly discuss the position of hydrogen in the
   Periodic Table.

   (c) Indicate how the oxides of the alkali metals may
   be used to illustrate the gradation in activity
   of a group of elements within the Periodic Table.

2. (a) Write out the product(s) of the following reactions,
   giving both the formula(e) and name (s):-

   (i) toluene $\xrightarrow{\text{nitration}}$

   (ii) nitrobenzene $\xrightarrow{\text{bromination}}$

   (iii) bromobenzene $\xrightarrow{\text{sulphonation}}$

   (iv) toluene $\xrightarrow{Br_2/uv \text{ light}/100^{\circ}C}$

   (v) 2-bromobutane $\xrightarrow{NH_3/S_N2}$

   (b) Write out the structures of important reaction
   intermediates involved in reactions (i), (iv) and (v)

3. (a) Write down the equilibrium constant for the reaction

$$N_2(g) + 3H_2(g) \rightleftharpoons 2NH_3 \qquad \Delta H = -92 \text{ kJ}$$
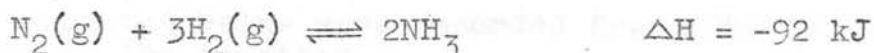
Describe what changes (if any) are brought about by

(i) adding some $N_2$ to an equilibrium mixture of $N_2$, $H_2$ and $NH_3$.

(ii) having <u>initially</u> present some $NH_3$ in addition to $N_2$ and $\overline{H_2}$.

(iii) increasing the pressure acting on the system at equilibrium.

(iv) increasing the temperature.

(b) At a certain temperature T, the value of $K_c$ for the above reaction as written is $0 \cdot 0001$. Into a vessel of 10 $dm^3$ capacity there is introduced varying amounts of the three gases $N_2$, $H_2$ and $NH_3$ as indicated below.

(i) $0 \cdot 1$ mole $N_2$ + 10 mole $H_2$ + 1 mole $NH_3$,

(ii) 1 mole $N_2$ + 100 mole $H_2$ + 1 mole $NH_3$,

(iii) $0 \cdot 01$ mole $N_2$ + 10 mole $H_2$ + 10 mole $NH_3$,

(iv) 1 mole $N_2$ + 10 mole $H_2$ + 10 mole $NH_3$.

For each of the above mixtures deduce whether or not any net reaction will occur and state the direction of any such reaction.

(c) The kinetics of the reaction

$$3A + B \longrightarrow 2C$$

was investigated and the rate law was found to be:
Rate of formation of C $= k [A]^2 [B]$
How is the rate of formation of C affected by each of the following changes ?

(i) Doubling the concentration of A

(ii) Doubling the concentration of B

(iii) Raising the temperature.

(iv) Adding a catalyst.

What is the overall order of the above reaction ?

How is the rate of formation of C related to the rate of disappearance of A and B ?

3. (d) The data tabulated below were recorded from a series of experiments on the reaction

$$A + B + C \longrightarrow D + E$$

| Experiment | Initial conc./mol.dm$^{-3}$ | | | Initial rate of formation of D / mol dm$^{-3}$ sec$^{-1}$ |
|------------|----------|----------|----------|------------|
| | $A_o$ | $B_o$ | $C_o$ | |
| 1 | 0·2 | 0·1 | 0·1 | 8 x 10$^{-5}$ |
| 2 | 0·4 | 0·1 | 0·1 | 32 x 10$^{-5}$ |
| 3 | 0·2 | 0·2 | 0·1 | 16 x 10$^{-5}$ |
| 4 | 0·2 | 0·1 | 0·3 | 8 x 10$^{-5}$ |

(i) Write down the rate equation.

(ii) Calculate the rate constant.

4. All the items in this question should be studied and where known the correct response to each item circled on the answer sheet. Candidates are advised not to guess as incorrect responses will lower the overall mark for this question. i.e. score = (Correct - $\frac{wrong}{3}$ ) response.

1. Under certain conditions an important side reaction in the conversion of t-butyl bromide to t-butanol with $OH^-$ ions is

    A.   formation of isobutene by elimination.

    B.   formation of hexamethylethane by Wurtz coupling.

    C.   isomerisation of t-butyl bromide.

    D.   stereochemical inversion of the t-butyl bromide.

2. Benzene reacts with ethyl chloride and aluminium chloride in nitrobenzene solvent to give mainly

    A.   isopropylbenzene.

    B.   naphthalene.

    C.   ethylbenzene.

    D.   m-diethylbenzene.

3. The orientation of the products (o-, m-, or p-) from aromatic substitution reactions is governed by

    A.   the nature of the reaction medium.

    B.   the type of substituent already present in the aromatic ring.

    C.   the relative stabilities of the products.

    D.   the delocalisation of the $\pi$-electron system.

4. t-Butyl bromide reacts with hydroxide ion by a first order process with the rate$\propto$[t-Bu Br]. One rationale for this is that

    A.   the intermediate t-butyl radical is stabilised by solvation.

    B.   the product t-butanol is thermodynamically stable.

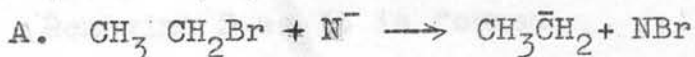    C.   in this case a stereochemical inversion is favourable.

    D.   formation of a very stable intermediate t-butyl carbonium ion is facile.
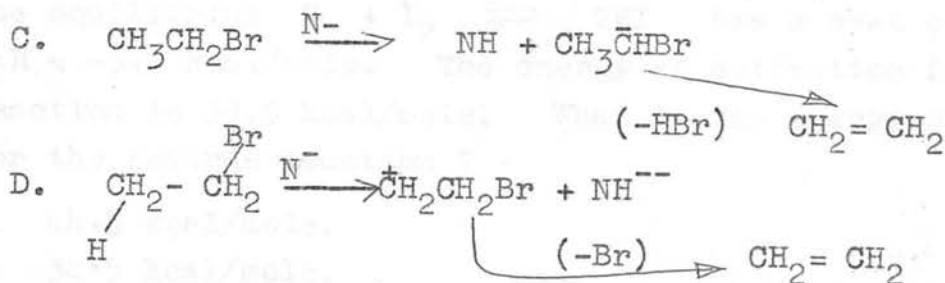
5. The mechanism of dehydrobromination of ethyl bromide by nucleophiles $(N^-)$ is best represented in the following way

    A.   $CH_3 CH_2Br + N^- \longrightarrow CH_3\bar{C}H_2 + NBr$

               $CH_3\bar{C}H_2 \longrightarrow CH_2= CH_2 + H^-$

    B.   $CH_2 - CH_2 \overset{Br}{\longrightarrow} NH + CH_2=CH_2 + Br^-$
                $|$
               H
             N

C. $CH_3CH_2Br \xrightarrow{N-} NH + CH_3\bar{C}HBr$

$\xrightarrow{(-HBr)} CH_2= CH_2$

D. $\underset{\overset{|}{H}}{CH_2- CH_2}$ with Br $\xrightarrow{N^-} \overset{+}{C}H_2CH_2Br + NH^{--}$

$\xrightarrow{(-Br)} CH_2= CH_2$
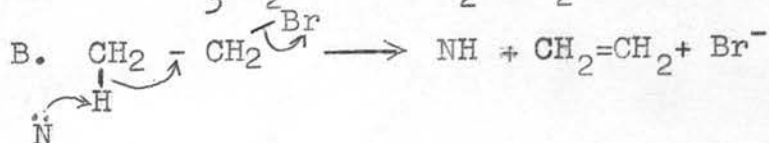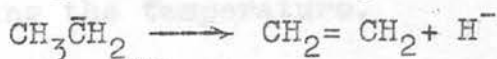
6. Which of the following is UNTRUE in relation to the molecule of benzene ?

A. All bond lengths are equal, and the molecule is planar.

B. All carbons are $sp^3$ hybridised, leading to resonance stabilisation.

C. The $\pi$-electrons are delocalised cyclically, both above and below the ring plane.

D. A complete picture of the molecule of benzene can be obtained by considering all resonance forms simultaneously.

7. For the reaction

$$2NO_2 \rightleftharpoons 2NO + O_2$$

the equilibrium constant $K_c$ if given by

A. $Kc = \dfrac{[2NO_2]^2}{[2NO]^2[O_2]}$

B. $K_c = \dfrac{[NO]^2[O_2]}{[NO_2]^2}$

C. $K_c = \dfrac{[NO_2]}{[NO][O_2]}$

D. $K_c = \dfrac{[NO_2]^2}{[NO]^2[O_2]}$

8. In the homogeneous gas reaction $W + X \rightleftharpoons Y + Z$ ($\Delta H$ positive) the yield of Y will be increased by

A. increasing the total pressure on the system.

B. Employing a suitable catalyst.

C. Removing Z as it is formed.

D. Lowering the temperature.

9. The equilibrium $H_2 + I_2 \rightleftharpoons 2HI$ has a heat of reaction $\Delta H = -5.0$ kcal/mole. The energy of activation for the forward reaction is 39.5 kcal/mole. What is the energy of activation for the reverse reaction ?

A. 44.5 kcal/mole.
B. 34.5 kcal/mole.
C. 49.5 kcal/mole.
D. 42.75 kcal/mole.

10. The effect of a catalyst in a reversible reaction is to
A. reduce the energy of activation.
B. add energy to the reaction.
C. lower the rate of the backward reaction.
D. alter the equilibrium composition.

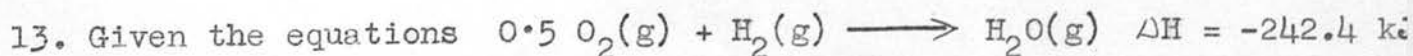11. A solution containing w grams of an organic compound in $Vm^3$ has an osmotic pressure of $\gamma$ $Nm^{-2}$ at T. K. The apparant molecule weight of the compound is

A. $\dfrac{w.RT}{V\,\gamma}$

B. $\dfrac{wT\,\gamma}{RV}$

C. $\dfrac{wT}{VR\,\gamma}$

D. $\dfrac{wR}{V\gamma T}$

12. $C + O_2 \longrightarrow CO_2$ $\qquad\qquad\qquad \Delta H = -x$ kJ
    $2CO + O_2 \longrightarrow 2CO_2$ $\qquad\qquad \Delta H = -y$ kJ

The heat (enthalpy) of formation of carbon monoxide is

A. $2x - y$

B. $\dfrac{y - 2x}{2}$

C. $y - 2x$

D. $\dfrac{y - x}{2}$

13. Given the equations $0.5\,O_2(g) + H_2(g) \longrightarrow H_2O(g)$ $\Delta H = -242.4$ k.

$\qquad\qquad\qquad 0.5\,O_2(g) + 2H(g) \longrightarrow H_2O(g)$ $\Delta H = -660.4$ k.

How much energy is needed to dissociate 1 mole of $H_2$
according to the equation

$$H_2(g) \longrightarrow 2H(g) ?$$

A. 242·4 kJ
B. 418·0 kJ
C. 660·4 kJ
D. 836·0 kJ

| | First Statement | Second Statement | Argument |
|---|---|---|---|
| A. | True | True | second statement is a correct explanation of the first |
| B. | True | True | second statement is not a correct explanation of the first |
| C. | True | False | |
| D. | False | True | |

14. Many covalent compounds are gases or volatile liquids

BECAUSE

covalent bonds are weak bonds.

15. Nitrogen gas is less reactive than other diatomic
gaseous elements

BECAUSE

the atoms in the nitrogen molecule are joined by a triple bond.

16. Magnesium is above calcium in the electrochemical series

BECAUSE

magnesium is above calcium in group two of the Periodic Table.

17. Hydrogen gas can oxidise some elements

BECAUSE

some elements will give up electrons to hydrogen gas.

18. Sodium hydroxide is a strong base

BECAUSE

sodium hydroxide is very soluble in water.

19. Sodium carbonate can be prepared by adding calcium
carbonate to sodium chloride solution

BECAUSE

calcium chloride is soluble in water.

20. Diamond is a very hard substance

## ASSERTION QUESTIONS

Two statement are made and are joined by the word <u>BECAUSE</u>.
Decide (1) if the <u>FIRST</u> statement is true or false, (2) if
the <u>SECOND</u> statement is true or false, and (3) whether the
second statement is a correct explanation of the first.   THEN
select one response from A-D according to the following table.

Summarised directions for assertion questions.

|   | First Statement | Second Statement | Argument |
|---|---|---|---|
| A. | True | True | second statement is a correct explanation of the first |
| B. | True | True | second statement is not a correct explanation of the first |
| C. | True | False | |
| D. | False | True | |

14.   Many covalent compounds are gases or volatile liquids
<center>BECAUSE</center>
covalent bonds are weak bonds.

15.   Nitrogen gas is less reactive than other diatomic
gaseous elements
<center>BECAUSE</center>
the atoms in the nitrogen molecule are joined by a triple bond.

16.   Magnesium is above calcium in the electrochemical series
<center>BECAUSE</center>
magnesium is above calcium in group two of the Periodic Table.

17.   Hydrogen gas can oxidise some elements
<center>BECAUSE</center>
some elements will give up electrons to hydrogen gas.

18.   Sodium hydroxide is a strong base
<center>BECAUSE</center>
sodium hydroxide is very soluble in water.

19.   Sodium carbonate can be prepared by adding calcium
carbonate to sodium chloride solution
<center>BECAUSE</center>
calcium chloride is soluble in water.

20.   Diamond is a very hard substance

# BECAUSE

the bonding in diamond is covalent and continuous
throughout the lattice.

## APPENDIX V

## The Post-Test

1. Respond to the following statement in one of the following categories.

   "As a means of instruction I prefer the conventional lecture to a programmed text".

   A.   Strongly agree

   B.   Agree

   C.   Undecided

   D.   Disagree

   E.   Strongly disagree

2. Write down the solubility product expressions for $BaSO_4$, $Ag_2CrO_4$ and $Ag_3PO_4$.

3. Arrange the following results in an order of decreasing solubility (indicate the decrease by an arrow)

   |          | $K_{SP}$           |
   |----------|--------------------|
   | AgCN     | $1.7 \times 10^{-16}$ |
   | $SrCrO_4$ | $3.3 \times 10^{-5}$  |
   | $AgIO_3$  | $3.0 \times 10^{-8}$  |
   | $BaCrO_4$ | $1.3 \times 10^{-10}$ |

4. The solubility of strontium carbonate $(SrCO_3)$ in water is $1.1 \times 10^{-3}$ g per 100 cm$^3$ of solution. What is the value of $K_{SP}$ for $SrCO_3$ ?

   $$M(SrCO_3) = 147.6 \text{ g mol}^{-1}$$

5. What is the solubility of AgCN in water ? Express your answer in mol dm$^{-3}$. $K_{SP}$ (AgCN) $= 1.7 \times 10^{-16}$

6. What is the solubility of $Ag_2CrO_4$ in water ?

   $$K_{SP}(Ag_2CrO_4) = 1.9 \times 10^{-12}$$

7. Aluminium hydroxide $Al(OH)_3$, has a $K_{SP}$ of $2.0 \times 10^{-32}$. Show how you would calculate the pH of a saturated solution of $Al(OH)_3$ ($K_w = 1.0 \times 10^{-14}$)

8. A solution is 1 mol dm$^{-3}$ in NaCN. Excess solid AgCN is added. What is the $Ag^+$ ?

$$K_{SP} \text{ (AgCN)} = 1.7 \times 10^{-16}$$

9. How many moles of PbBR ($K_{SP} = 5.7 \times 10^{-6}$) will dissolve in a 0.01 mol dm$^{-3}$ NaBr solution ?

10. A solution is 0.01 mol dm$^{-3}$ in both Ferrous ($Fe^{2+}$) and cupric ion ($Cu^{2+}$). Sulphide ion is gradually added.

(a) Will FeS ($K_{SP} = 1.0 \times 10^{-19}$) or CuS ($K_{SP} = 1.0 \times 10^{-35}$) precipitate first ?

(b) What will be the $\left[ Cu^{2+} \right]$ when FeS just starts to precipitate ?

## APPENDIX VI

## Data for Experiment 2

| Group A (Lecture) | | Group B (Programme) | |
| Pre-Test Score $(x)$ | Post-Test Score $(y)$ | Pre-Test Score $(x)$ | Post-Test Score $(y)$ |
| --- | --- | --- | --- |
| 60 | 50 | 55 | 47 |
| 56 | 65 | 68 | 74 |
| 32 | 53 | 72 | 82 |
| 51 | 79 | 66 | 88 |
| 54 | 50 | 50 | 79 |
| 71 | 74 | 46 | 59 |
| 54 | 71 | 60 | 68 |
| 44 | 79 | 50 | 74 |
| 64 | 71 | 57 | 82 |
| 35 | 82 | 48 | 94 |
| 75 | 77 | 54 | 21 |
| 47 | 71 | 54 | 82 |
| 55 | 44 | 52 | 74 |
| 48 | 21 | 75 | 71 |
| 26 | 0 | 49 | 62 |
| 51 | 77 | 45 | 32 |
| 25 | 24 | 30 | 68 |
| 71 | 77 | 47 | 56 |
| 46 | 62 | 25 | 47 |
| 57 | 85 | 50 | 77 |
| 58 | 74 | 65 | 79 |
| 40 | 68 | 69 | 91 |
| 73 | 94 | 30 | 79 |
| 51 | 79 | 39 | 94 |
| 54 | 47 | 52 | 50 |
| 44 | 71 | 51 | 41 |
| 61 | 65 | 60 | 53 |
| 44 | 59 | 57 | 77 |
| 73 | 91 | 52 | 50 |
| 65 | 62 | 50 | 50 |
| 71 | 94 | 45 | 53 |
| 76 | 91 | 77 | 91 |
| 70 | 77 | 46 | 18 |
| 40 | 35 | 42 | 85 |
| 70 | 91 | 73 | 77 |
| 50 | 35 | 65 | 82 |
| 48 | 65 | 57 | 100 |
| 46 | 77 | 53 | 71 |
| 63 | 88 | 22 | 59 |
| 49 | 62 | 23 | 82 |
| 31 | 35 | 66 | 74 |
| 44 | 59 | 74 | 94 |
| 65 | 74 | 64 | 74 |
| 44 | 74 | 67 | 82 |
| 83 | 77 | 56 | 65 |
| 62 | 44 | 49 | 82 |
| 56 | 91 | 59 | 41 |
| 59 | 68 | 42 | 94 |