The UK Higher Education Academy Physical Sciences Centre

ISBN 1-903815-07-X

Getting Started in Pedagogical Research in the Physical Sciences

A Physical Sciences Practice Guide

Norman Reid

October 2006

UK Physical Sciences Centre
Department of Chemistry
University of Hull
Hull HU6 7RX
Phone: 01482 465418/465453 Fax: 01482 465418
Email: psc@hull.ac.uk Web: www.heacademy.ac.uk/physsci

Getting Started in Pedagogical Research

in the

Physical Sciences

Norman Reid
Centre for Science Education
University of Glasgow

Published by the Higher Education Academy UK Physical Sciences Centre

Getting Started in Pedagogical Research in the Physical Sciences

Purpose of this text

This booklet is written for the many lecturers in the physical sciences who have a concern for and interest in the teaching of their disciplines. It is a challenging task to share something of what we know with future generations to enable them to discover the excitement of exploring the physical world.

In the annotated bibliography [available at <www.heacademy.ac.uk/physsci/publications/practiceguides>], an attempt has been made to summarise the main findings from pedagogical research in the physical sciences. This text seeks to introduce the ways in which our understandings of teaching and learning can be explored.

It is hoped that readers will find this booklet useful as a guide and pattern as they begin their own pedagogical research.

The author gratefully acknowledges the contributions of many. Particular thanks must go to Tina Overton who encouraged me to take on the task of compiling this booklet. I should also like to thank Professor Alex H Johnstone and Professor Rex Whitehead who made most useful comments on an earlier draft.

This version printed June 2010.

Norman Reid
Centre for Science Education
University of Glasgow
July 2003

Getting Started in Pedagogical Research in the Physical Sciences

Contents

Introduction

As university teachers, we usually give lectures, lead tutorials and supervise laboratories. We carry out our research and supervise our research students.

We are committed to our subject specialities and are encouraged when our students share a little of that enthusiasm and competence. However, most of us would admit to the many times when we find the progress of our students frustrating and wonder if there is anything we can do to make success more easily attainable for them. We may even be concerned about declining numbers and our abilities to retain our students in our courses.

How do we find out ways by which learning can be made more effective? How can we find out how to attract and hold our students? This is the focus of this booklet.

Pedagogical research simply means finding out more about how learning takes place so that we, as teachers, can direct our energies into approaches which are more likely to be successful.

How to Use this Booklet

Try reading pages 1 to 18 to gain an overview of pedagogical research.

In the light of your own interests, you might wish then to move to specific themes:

For written tests, go to page 19
For interviews, go to page 28
For how to measure attitudes, go to page 32

There are also sections which deal with specific topics and themes:

Data handling and statistics - pages 42-50
Drawing conclusions - pages 51-53
Aiming at quality in research - pages 54-56

Pages 57-61 offer summaries of some examples of pedagogical research which give examples of useful approaches.

The best way to learn how to conduct pedagogical research is to carry out pedagogical research. Try things out for yourself!

Pedagogical Research

Asking Questions

This section seeks to describe the nature of pedagogical research.
It stresses that the focus is on teaching and learning in the physical sciences.
It outlines the likely starting points for such research.

Research means asking questions. This implies that:

There are questions worth asking.
There are some general principles which will assist effective and efficient learning.
We believe in the possibility of improvement.

Research can be defined as any systematic, critical and self-critical enquiry which aims to contribute to the advancement of knowledge. Education in the physical sciences can be thought of as involving the acquisition of understandings, attitudes and skills.

This booklet will ***not*** tell you whether 'chalk and talk' is better than using 'PowerPoint' - that question is not worth asking. The answer might depend on what you are teaching, who you are, what your resources are, the numbers of your students, your aims and goals and so on. Socrates used neither!

This booklet will ***not*** develop abstract theories of learning. These may have their place but not here.

It will seek to give you the ***basic strategies and approaches*** which might allow you to find out more about teaching and learning, how the human mind assimilates and masters new knowledge and experiences in meaningful ways and how student attitudes and perceptions can be considered as a helpful guide to future planning.

Pedagogical research allows us to be ***scientists with regard to our teaching*** in ways similar to the approaches we adopt in our research in the physical sciences. We are professionals with regard to our teaching and not merely university functionaries. Academic educational research too often lives in a different intellectual world from that of the practitioner lecturer and so the two areas rarely connect. This booklet seeks to bring pedagogical research and the practitioner closer together.

Some Wrong Assumptions

(1) **Teaching and Learning are All matters of Opinion**: There are natural differences between people and this can affect the way different students prefer to learn and the different style adopted by different lecturers. For example, some students prefer to learn by means of visual input (diagrams, pictures, graphs) while others prefer to learn by using symbols (language or mathematics). Most can learn using both approaches. Similarly, some lecturers like using 'PowerPoint' type approaches while others use the overhead projector or blackboard extensively.

Despite these differences between people, there are ***general principles underpinning the way all of us learn***. If we teach in line with these general principles, then learning will be maximised. If we teach in a way which is inconsistent with these learning principles, then learning will be reduced. The general principles were once summarised thus:

1. What you learn is controlled by what you already know and understand.
2. How you learn is controlled by how you have learned successfully in the past.
3. If learning is to be meaningful it has to link on to existing knowledge and skills enriching and extending both.
4. The amount of material to be processed in unit time is limited.
5. Feedback and reassurance are necessary for comfortable learning and assessment should be humane.
6. Cognisance should be taken of learning style and motivation.
7. Students should consolidate their learning by asking themselves about what is going on in their own heads.
8. There should be room for problem solving in its fullest sense to exercise and strengthen linkages.
9. There should be room to create, defend, try out, and hypothesise.
10. There should be opportunity given to teach (you don't really learn till you teach).

(after Johnstone, 1997[1])

(2) **Knowing Your Subject makes you a Good Teacher**: This is simply ***not*** true. Knowing is no guarantee of any ability to be able to communicate effectively. However, if the teacher does not know his/her subject, then students will learn only by finding some other way to the knowledge.

(3) **Experience Makes a Good Teacher**: This may or may not be true. We have all met highly experienced lecturers who still leave their students relatively uneducated. If we sensitively learn from experience and evidence, then excellent teachers can become brilliant, good teachers can become better and even poor teachers can improve. The "if" clause is important.

(4) **Students are "Empty Pots to be Filled"**: All the evidence shows that this is not true. Students are known to construct their own knowledge and understanding for themselves. What we offer them will not be what they receive and make their own, as all markers of examination scripts are aware. The lecturer's task is not to transmit knowledge from one head to another. The task is to allow the students to develop their own knowledge and understanding and make sense of the material being taught.

---

[1] Johnstone, A.H., (1997) Chemistry Teaching, Science or Alchemy?, Journal of Chemical Education, **74** (3) 262-268.

What is Pedagogical Research ?

Research is about finding out, generalising and establishing broad principles. The early work in science education research started by finding out what difficulties students have in learning chemistry and physics. Many of these areas of difficulty were then explored in detail to try to find out what it was that was the source of the problem. Eventually, patterns began to emerge and these developed into broad principles.

If you want to find out something of the findings in relation to teaching and learning in the physical sciences in higher education, go to:
<www.heacademy.ac.uk/physsci/publications/practiceguides>

Astronomy, Chemistry and Physics

We would not dream of basing the astronomy, chemistry or physics we teach on our opinions and views. We base our understandings of these disciplines on established concepts and principles which have been derived from myriads of experiments, conducted over many years.

Teaching and learning are exactly the same. Just as abstract ideas and mathematical formulations are used to describe our understanding of the physical world, pedagogical research has developed abstract ideas to describe the processes of teaching and learning. We cannot see the electron but can describe its behaviour in many useful ways. Similarly, we cannot see the development of a concept in the human brain but we can describe in many ways the general 'rules' for the way concepts develop. Such rules can be useful to us as teachers.

We address our questions to the physical world in astronomy, chemistry or physics by means of all kinds of physical measurements, complex computer models and mathematical manipulations. We address our questions to the learner in pedagogical research by means of observation, evidence gathered in a written form, measurement of student performance in relevant tasks. We may apply computer analysis to the data we obtain in an attempt to generalise and establish the underlying principles.

In looking at the physical world, we select and simplify in order to be able to deduce general principles. We do this almost automatically. We are very adept at using resistance-free wires, frictionless pulleys, concentrations as approximations to activities, and so on. Indeed, one of the arts of understanding the physical world is to be able to select those variables which are of interest, ignoring, randomising or reducing the impact of other variables. Exactly the same is true of pedagogical research. We need to learn, by experience, how to control variables, how to randomise, how to select, how to simplify to make meaningful deductions.

There is one important complication. We can safely assume for most purposes that particles like electrons are all identical. The same is never true of our raw material (our students) in pedagogical research. No student is intrinsically identical to any other student. Thus, to draw meaningful conclusions, we need to have large numbers of students, to randomise all the variables which define the differences between one student and the next.

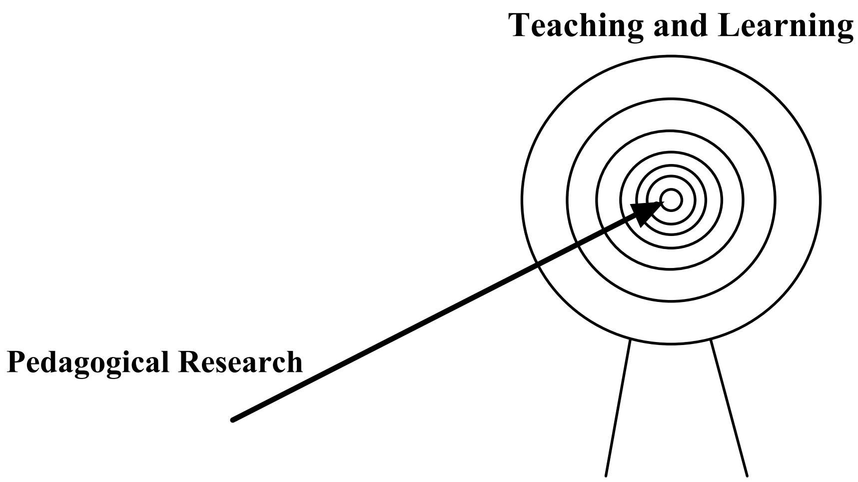
Why Pedagogical Research?

Pedagogical Research in likely to involve:

- Experiments as sources of evidence
- Models as useful ways to summarise
- Data gathering by means of observation and testing
- Practical feasibility by selecting and simplification
- Controlling variables by selecting and randomising

Most of us who are involved in teaching in higher education have a commitment to our own discipline and subject specialisms. With that comes a desire to see our students develop competence, confidence and enthusiasm.

Pedagogical research focusses ***relentlessly*** on the teaching and learning process. Pedagogical research must not become utopian and out of touch with the classroom, lecture theatre or laboratory nor must it depend on statistical approaches where statistics appear to prove anything you like. Pedagogical research is at its best when it assesses current practice, justifies good practice, looks in detail at teaching and seeks to find out how students actually learn successfully. This may allow us to develop teaching and learning characterised by improved competence, confidence and enthusiasm.



In many continental countries, the concept of pedagogy is held in much higher respect than here. Perhaps in the UK, there has developed an unreal and unhealthy divide between the teacher-practitioner and those who think about learning. We need to restore an emphasis on what students experience during the learning process. The focus must be on the student learner. Teaching and learning are ***not*** all matters of opinion nor are they the province of the theoreticians ! "...the swampy lowlands of practice and the elevated highlands of theory..." (Schon, 1983)

Models for Research

There is a problem in how we view pedagogical research. Much pedagogical research has tended to adopt a model of research derived from the worlds of agricultural or clinical trials where variables are controlled and there are matched control and experimental situations.

Compared to agriculture, there are some difficulties in that, in teaching and learning, it is not easy to control variables and contexts. Indeed, one great difficulty is that performance criteria are not as simple as better crop yield! Teaching and learning involve interactions between teacher and learner as well as between learners. All this is part of the learning process but it is difficult to control. Indeed, the teacher treats different learners in different ways and different learners may prefer to learn in different ways. Ideally, the teacher is like a gardener who treats each plant differently.

This means that pedagogical research must be conducted with care and conclusions drawn with caution. It also makes it difficult to draw overall conclusions which apply in every circumstance. It is too easy to find pedagogical research which is trivial in the sense that it does not take our understanding forward in any meaningful way, no matter how carefully it has been conducted. Nonetheless, over the years, some research has generated general patterns which are widely applicable and our research today must build on these.

How do I get Started?

- **Do you have Burning Questions?**

It is possible that you have burning questions which you wish to explore and your problem is to know how to find answers. It is difficult to do research effectively if you do not have some questions whose answers you feel you must know.

Three examples, based on some real projects, illustrate this way forward:

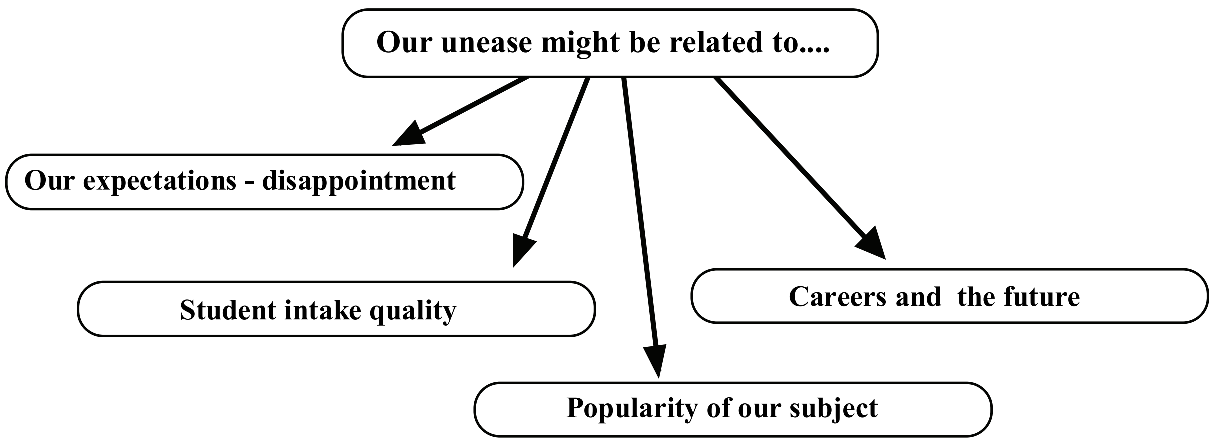
(i) One lecturer observed that some students in an honours course in flow mechanics seemed to like highly mathematical approaches while others found a more visual approach helpful. This gave rise to a whole series of questions: are some students, by nature, orientated towards the more visual approach? Given a free choice, what proportions of students would opt for each way of learning? Are there other factors which influence the choice of approach (like examination styles, student expectations)?

(ii) One lecturer observed that tutorial groups tended to be just like mini-lectures. Students were not interacting with the materials offered in a meaningful way but relied too heavily on lecturer input. Student-student interactions seemed to be rare. This led to an exploration of student attitudes to learning, their view of the role of the lecturer, and the development of new learning materials which moved the learning process away from the lecturer to the student in the form of group work problem solving.

(iii) One lecturer was sceptical of the efficiency of lecturing as a teaching approach. At the end of a lecture (which had been recorded), students were invited to have their lecture notes photocopied. These photocopies were analysed to find out what they had recorded in their notes and these records were compared to what the lecturer had offered. This offered quite amazing insights into the limits of student notes and the variety of emphases students had discerned. It offered clear directions to the lecturer about what he could do to improve student learning.

- **Do you have General Unease?**

For many involved as teachers in the physical sciences, the pedagogical problems are not so clear cut. There may be general feeling of unease, that things could be done better, that student performance is not matching expectation. There are several general approaches which can be adopted:



(a) Disappointment

For many of us, we may have had the sad experience of marking examination papers based on a course we have taught on a topic which is important and interesting to us. Our student scripts have fallen far short of our expectations! Our reaction may vary from blaming our students (and the remunerative part-time jobs so many seem to have) and their lack of commitment to despair about our teaching.

Where do we start? Here are some ways which have been found to be useful.

(a) Carry out a difficulties survey, following a method widely used elsewhere[2]. This is easy to design, apply and mark and is known to give useful patterns.
(b) Look at those topics which students find most difficult (as detected by our difficulties survey) to see if there are common features;
(c) By means of questionnaire and interview, explore with students why these specific topics are causing problems.

(b) Popularity

In many parts of the world, the retention rate in the physical sciences is not high. School students may be opting out of the physical sciences and may be dropping out of our courses.

Where do we start? Here are some ways of clearing our minds about this.

(a) First of all, recognise that there is very little we can do about school experience. Past research has shown clearly the factors which influence school pupils towards the physical sciences. Almost none of these are open to modification by a university lecturer. Displays and shows run by universities, talks from university lecturers and visits by school students to universities have marginal impact.

(b) At the outset, we need some evidence of the extent of the problem. By looking through departmental data for many years, we can establish whether there is any pattern of loss. Are rates of loss growing? Are numbers opting for our subject falling? At least 10 years is needed to see a clear pattern.

(c) It is possible to explore our own students' attitudes towards our own subjects and the way our courses are offered. This requires the development of very carefully constructed questionnaires and the conduct of carefully planned interviews. These together can give a vivid picture of the patterns of students' attitudes and may lead to considerable further research.

(c) Intake

Frequently, lecturers comment on the quality of student intake from the schools, often negatively.

Where do we start? Here are some ways which have been found to be useful.

(a) Some apparently simple data gathering is often a useful starting point. Most university departments have records going back years and these show entry qualifications and numbers enrolling. An analysis of such data can reveal whether entry qualifications have been changing and, indeed, whether the range of school subjects offered by students on entry has shown any pattern or trend.

---

2 Johnstone, A.H., Morrison, T.I., Sharp, D.W.A., (1971) Topic Difficulties in Chemistry, Education in Chemistry, 8, 212.

(b) Contact with local schools to talk to appropriate staff may throw further light upon the type of student gaining entry over a period of time. It may also be possible to gain access to the school syllabuses which are current and those which were used, say, ten years before. Perhaps the school syllabus emphasis has changed in key ways which have not been allowed for in our courses.

(c) All of this may identify specific areas of our curriculum which need attention. It may point up the need to develop support materials, to modify teaching approaches, or to change the emphases of our subject matter. It may be possible to test out our new ideas with senior school pupils, collaborating with the schools in our area, to see if more confidence and competence is gained.

(d) Careers and the Future

For many students in the physical sciences, a career working in the practice of that science is often seen as the goal after graduation. However, the statistics of employment show that large numbers of our gradates do ***not*** pursue that path but follow occupations which are not dependent specifically on the content of astronomy, chemistry or physics. We may have observed this and wondered if we are serving our students well or we may fall back on the 'trained minds' argument.

Where do we start? Here are some ways which have been found to be useful

(a) To find out what students need apart from the content of our disciplines, we need to approach employers or students in recent employment. Sadly, we tend to gain a different impression from each employer we ask and each student we talk to! To gain a meaningful picture, we need to talk to hundreds of employers and thousands of students and this is simply not practicable. However, this has been done by others[3]. Harvey *et al*. has shown the key areas where employers see 'gaps'. How do we attempt to fill these?

(b) Many refer to generic skills of graduates but there is little analysis in the literature to establish whether such skills are generic or context-specific. Suppose we want to develop communication skills (which may possibly be generic). We have to define what kind of communication skills we have in mind: let us say, verbal communication by means of a mini-lecture. How can we judge performance fairly? If we do not award marks, the students will not take it seriously. We now have a nice topic for research enquiry.

(c) We can have, say, 6 assessors, working with a tight specification which we design, mark a set of, say, 20 students giving a 15 minute presentation of an appropriate topic - maybe a summary of research evidence underpinning some topic. We then analyse statistically the reliability of our markers, looking for those areas where the markers found difficulty. From this, we can develop step by step a procedure which is time efficient but reliable for marking some verbal presentations as an integral part of our course.

(d) That is but the starting point. The real question is how best to develop such skills in our students, thus preparing them better for the workplace. Is it just a matter of practice? Are some just born better than others? Does feedback assist and, if so, what kind is most effective? A myriad of research projects is now evident.

---

[3] Harvey, L and Green, D. (1994), Employer Satisfaction, *Quality in Higher Education Project:* University of Central England, Birmingham. Harvey, L., Moon, S. and Geall, V. (1997), Graduates' Work: Organisational Change and Student Attributes, University of Central England, Birmingham.

How Can we Know ?

This section looks at sources of evidence and the general way we conduct pedagogical research.

It considers the place of educational models underpinning research.

Sources of Evidence

In astronomy, chemistry and physics, we seek answers to questions by carrying out experiments. If we want to measure the pH of a solution, we may choose to use pH paper or a pH meter. Voltage might be measured by a meter, a potentiometer or oscilloscope. We select the instrument appropriate for the task and, perhaps, the size of our budget. It is exactly the same when we look at teaching and learning. We need to decide what we want to measure and then select an appropriate instrument to make the measurement.

The more fundamental question is to ask at the outset what do we want to find out. In addition, is it worth finding out? There are some questions which any experienced university lecturer will know without carrying out an experiment. For example, it is well known to lecturers involved in inorganic and analytical chemistry (and others) that calculations involving moles cause difficulties. There is no point in seeking to find out if our students have difficulties in such areas. It is much more worthwhile to seek to explore what are the underlying reasons WHY students consistently find the difficulties. In everything, we need to ask if the questions we are asking are likely to lead to answers which will enable students to learn better.

Teaching and Learning

What do want to know ?
Is it worth knowing ?
What is useful to measure ?
Have we a way to make the measurement ?
Will the measurement tell us anything useful ?

Pointless Research

We also have to ask if our chosen measurement technique will tell us anything useful. For example, if we wish to explore the reasons why a topic like thermodynamics is difficult, then using a multiple choice test is likely ***not*** to be too helpful, simply because there is no way of using such a test to give insights into WHY the student selects the wrong answers. Indeed, traditional multiple choice tests rarely offer useful information about the learning process.

In the same way that there are numerous aspects of the physical world which we think are important to explore and to measure, there are numerous aspects of the process of learning which we may wish to explore and to measure. In the same way that we usually try to check our measurements of the physical world, we also need to check our measuring of the world of learning.

How do we do this ?

Physical World	World of Teaching and Learning
Calibrate a measuring instrument	Check a test against a known situation
Repeat measurements	Repeat measurements
Specify error limits	Specify level of certainty
Look for potential errors	Look for potential errors
Check different methods for consistency	Check different methods for consistency
Simplify reality to make measurement possible	Simplify reality to make measurement possible
Check control of variables	Check control of variables
Develop hypotheses	Develop hypotheses

We need to know if we are measuring what we think we are measuring. This is not always easy in the physical world but it is even more difficult in the world of learning.

The Ideas of Validity and Reliability

In the physical world, we need to know the reliability of our measuring instruments. Of course, we cannot measure microamps accurately with an ammeter, nor can we titrate using measuring cylinders. However, the problem is more demanding than that. If we are using a burette, we need to have some idea that the burette will work reliably from day to day and with different experimenters. Is our photocell able to give similar results day after day, without being over-sensitive to changes in external conditions. In other words, we may need to know if our measurements are reliable in the sense of being able to be *repeated* with similar results.

We also need to know that what we observe is what we think we are observing. The sciences are littered with examples where this did not happen. The debacle of cold fusion is one such example. Another is the story of polywater.

In pedagogical research, it is not easy to be sure that what we are measuring is what we think we are measuring. In a recent project, a group of students were given a diagnostic test in coordinate systems in astronomy. These students could pass the examinations but the diagnostic test revealed major areas of conceptual weakness. It would seem that they were passing the examinations by means of recall of algorithms and skill in 'plugging in' numbers rather than on the basis of sound understanding. Thus, the final examination was not testing understanding but reflecting abilities in recall and applying routine procedures.

If we are fairly sure that what we are measuring is what we intend to measure, we describe our test as valid. The simplest way to check the validity of our testing is to check the results against some external criterion. For example, if we are trying to measure attitudes to physics and to explore retention rates, then a check of those who have continued with physics will confirm if our measurements are valid. Similarly, if we want to check if a test in chemistry measures understanding of key concepts, then one way would be to discuss our test with a group of colleagues and then try it out with a small group of students, talking to them as they complete the test.

For more details of validity and reliability, go to pages 51-52.

Controlling variables

One of the great problems with all experimentation lies in the need to control variables. Pressure-volume relationships cannot be established without control over the temperature and amount of material present. Even synthetic organic chemistry becomes impossible if we do not take into account the possibilities of solvent-reactant interactions as well as reactant-reactant interactions

Some measurements are relative. 40˚C is 20 Celsius degrees higher than 20˚C. However, some measurements have to be related to a baseline or zero point on a scale. Using degrees Celsius in thermodynamics is unhelpful in that the temperatures must be related to absolute zero on the scale. There are many chemistry undergraduates who have learned this the hard way when trying to plot graphs using the Celsius scale when exploring kinetics experiments only to obtain meaningless curves. For more information, go to page 16.

Hypothesis Formation

Some experimental work in the physical sciences is of a survey kind where there is an attempt to observe a new area of interest and look for patterns. Often, there is an absence of pre-conceived ideas although avoiding preconceptions is extraordinarily difficult. However, most experimental work carried out in physical science research starts with some underlying hypothesis or idea in mind. We make mental or verbal statements like, "If we can measure X, then it will be possible to deduce Y or conclude Z." We may be confirming ideas already established and extending them or, perhaps, we may be seeking to develop or refine ideas. Occasionally, we develop an experiment which is critical in the sense that its outcomes make it very clear that one understanding is likely to be correct and other understandings are wrong. The place of hypothesis formation and critical experiments is also important in looking at learning.

A series of experiments illustrates this.

> At the start, in a study of the way students 'see' organic structures and relate this to their difficulties, it seemed that the amount of information a student faced at any one time was important. In a major experiment which followed, students were set questions in chemistry. The questions varied in the amount of information a student had to hold mentally at any one time in order to reach an answer. This was estimated by a group of experienced teachers. The hypothesis was that the greater the information load, the poorer would be the performance. The results showed that this was generally true. However, a careful look at the relationship between student success and information load also showed very clearly that the performance dropped ***catastrophically*** when the load reached 6. This was not expected and led to the development of a quantitative way of understanding difficulty in the learning of the sciences.

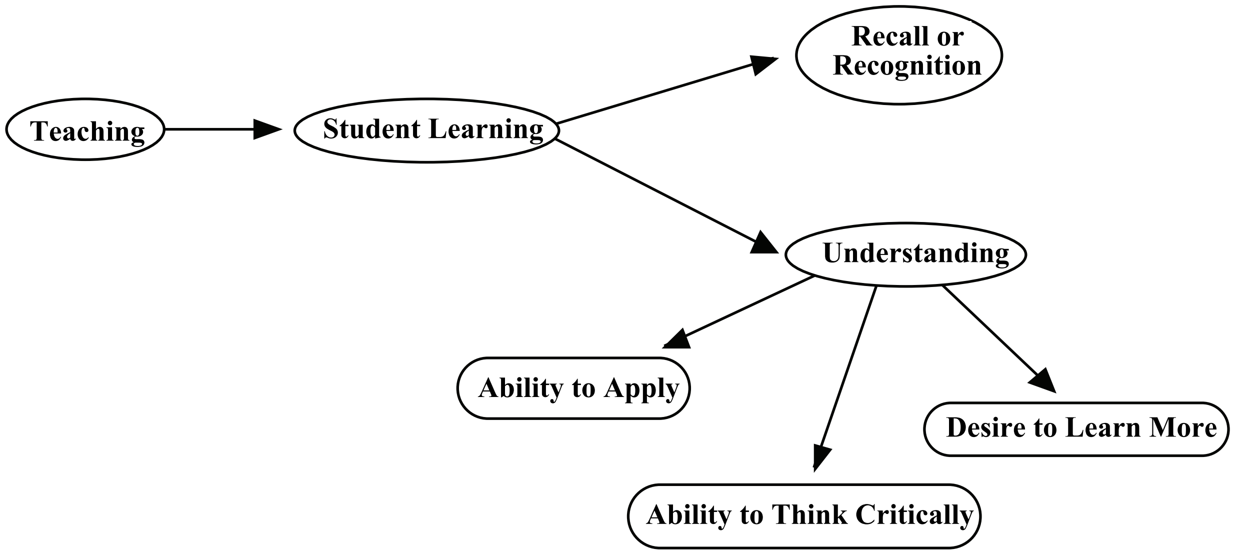
This experiment turned out to be critical and its results altered the ways of thinking about difficulty. It was later repeated in many countries in many contexts with many different kinds of students, always with the same general pattern of outcomes. [If you want to know more abut this work you might like to read: "Understanding Learning Difficulties - A Predictive Research Model", Studies in Higher Education, 14(2), 1989, 159-68.]

Types of Evidence

In essence, if we want to know about learning, then we have to test what students have actually learned. However, it is not quite that simple. Most of us remember scoring well in some area of our undergraduate studies and then appreciating, when we came to teach it, that we did not really understand it. Many tests and examinations test recall or recognition. They may do this very efficiently and effectively. However, recall is no guarantee of understanding.

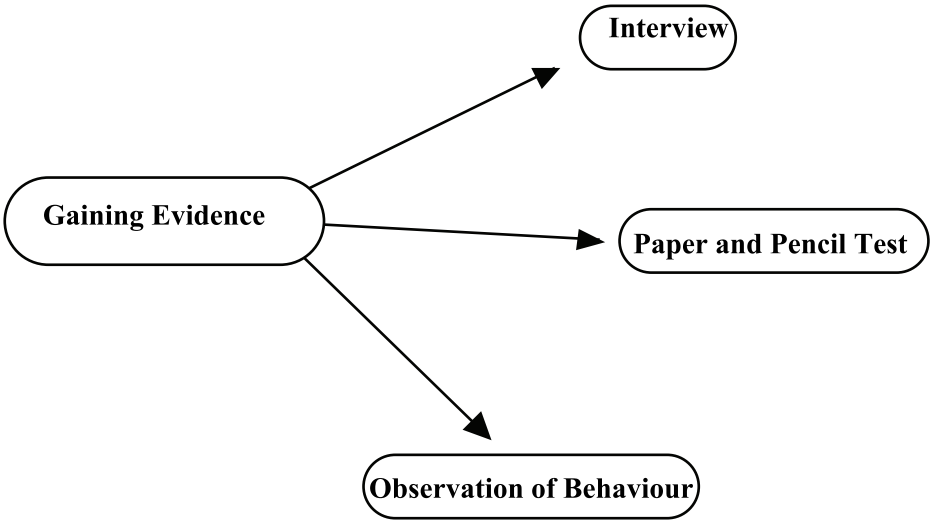
If we want to find out about student recall, then any test demanding recall will do.
If we want to test student understanding, then testing becomes much more difficult.

Suppose we re-cast our lecture course with the aim to increase student understanding. We now want to test if our students have, in fact, understood what we have taught (and are not just remembering it), then we need to develop a test designed in such a way that success can ***only*** be achieved by those who genuinely understand the lecture course. Students are clever and they develop all kinds of devious ways to pass tests and examinations. They do this to avoid the great cognitive effort involved in really understanding what is taught!



It is possible to think of the ways to gain evidence about learning by three main routes: interviews, paper and pencil tests and observation of behaviour. Interviews can be conducted individually or in small groups but, either way, they are time consuming although they can offer rich insights.

Three Ways to Gain Evidence



The range of possibilities for paper and pencil tests is simply enormous, while observation of behaviour can range from observation of the way students react in a lecture theatre to the controlled observation of practical skills.

If you want to know more about interviews, see pages 28-31.
If you want to know more about paper and pencil tests, see pages 19-27
If you want to know more about the observation of attitudes and related areas, see pages 32-41.

The Place of Theories

A theory can be described as a general statement or framework that summarises and organises knowledge by proposing a general relationship between events. However, sometimes the word is used merely to summarise a range of evidence. Yet, theory has to explain, to offer a rationalisation for observations, to offer the language and space to make analysis possible. A good theory offers prediction which is testable.

The word 'theory' has several meanings and some carry unfortunate overtones.

(a) The word theory can mean the opposite of practice;
(b) Theory can mean a formal hypothesis which we are using;
(c) Theory can be used as a kind of developing explanation;
(d) The word is sometimes used to mean some kind of overarching scientific model or principle.

The word 'theory' carries similar unfortunate overtones and shades of meaning in pedagogical research. It is often used in a derogatory sense, implying that the real thing is 'practice'. Indeed, it must be admitted that some educational theories have not been applicable to real learning situations. However, the word tends to be used in the sense of a developing explanation. On the basis of many observations and measurements, the researcher develops a kind of mental model (often expressed in diagram form) to rationalise what is observed.

There is general agreement that theories in educational research can do many things:

(a) They can provide explanation of observations;
(b) They can offer interpretations of observations;
(c) They can allow findings to be generalised;
(d) They can have offer logical consistency;
(e) They can facilitate prediction;
(e) They are provisional.

Overall, a theory can be seen as containing a network of interrelated generalisations or laws.

Theories in the physical sciences.

In the physical sciences, theoretical concepts are important. Ideas as diverse as hybridisation, orbital splitting, the equation of time and the conservation of momentum are attempts to summarise and rationalise experimental observations. In the same way, pedagogical research has underpinning models such as information processing, the law of planned action and cognitive conflict. In the physical sciences, there are constructs such as mass, force, energy, proton, photon while pedagogical constructs include working memory space, creativity, self-concept, and visual thinking. Constructs are not directly observed but inferred from behaviour (physical or human).

Some parts of pedagogical research can be seen as parallel to scientific research. The sciences can be thought of as organised and tested knowledge of natural phenomena while pedagogical research can be thought of as organised and tested knowledge of the ways learning takes place.

Controlling Variables

As in the physical sciences, there is often difficulty in controlling variables. While the argument that the unity of all science consists only in its method is largely discredited, patterns of scientific thinking comprise a range of related approaches. The neat sequence of procedures (problem, hypothesis development, hypothesis testing by experiment, conclusion drawing) is not often followed in physical science research. There may be all kinds of observation stages.

Similarly, in pedagogical research there are many approaches. Basically, if we want to know something, then we have to go out and have a look. Sometimes, we can use methods which are very similar to those used in the sciences, medicine or the social sciences. However, more qualitative, holistic observation and dialogue may be the better way to approach some questions.

The more scientific way of approaching pedagogical research involves careful observation, systematic recording, attempts to generalise findings to develop models of understanding and then the use of such models to predict. One classic example in the UK is Johnstone's work. He observed what were the difficulties in learning chemistry, he then explored many of these difficulties in detail to try to find out what was causing the difficulty and to find common factors. He then developed a model which seemed to account for all the observations. He then used the model to predict what would happen if changes in teaching took place and he found that the predicted outcomes were observed and his model became steadily more established.

Another example arose from studies in attitudes. A huge literature survey in social psychology sought to pinpoint key strategies which might allow attitude development in an educational setting; A hypothesis was developed from this and was tested in action. The outcomes supported the usefulness of the hypothesis. Further studies on attitudes to learning were carried out and the outcomes also supported the hypothesis. For more details of these examples, go to:
<www.heacademy.ac.uk/physsci/publications/practiceguides>

Nonetheless, the scientific approaches must be handled with care. We are not as objective as we sometimes think at the observation stage. Observations are filtered through the understandings, preferences and beliefs (maybe prejudices!) of the observer. The experimental stage is fraught with difficulties in controlling variables. Indeed, the effect being sought can be too small to observe amid the other effects and the actual presence of some intervention may have effects of its own.

All this emphasises the importance of applying appropriate care and common sense, talking to colleagues (as critical friends) and looking always for corroborative evidence to support any findings we make. The world of scientific research pursues exactly the same strategies in its attempt to be rigorous.

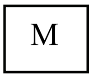
Carrying out Pedagogical Experiments

In the research areas encompassed by the physical sciences, we depend on *research evidence* to draw conclusions, to develop hypotheses and to generate models to aid our understanding of the physical world. This research evidence is gained mainly by means of experiments and calculations.

It is exactly the same with the world of teaching and learning. We need to conduct well designed experiments, we need careful analysis of the data obtained and then we need to draw conclusions based on the evidence. Teaching and learning is no more the province of opinion and assertion than our own specialist areas of research in the physical sciences!

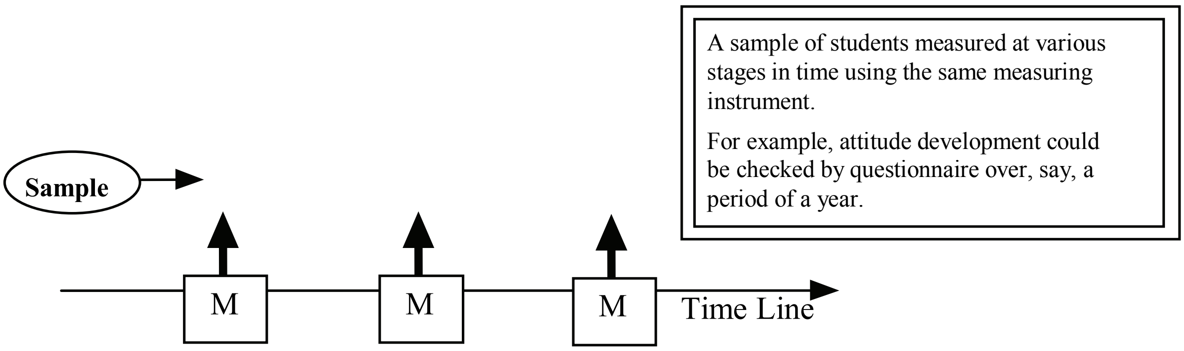
However, pedagogical research has an added complication - the control of variables is often much more difficult in that there are so many teacher and learner variables. Therefore, we often need to consider looking at large samples of students and teachers to gain a meaningful overall picture.

Here are examples of five possible ways to start thinking about setting up experiments to gain our data.

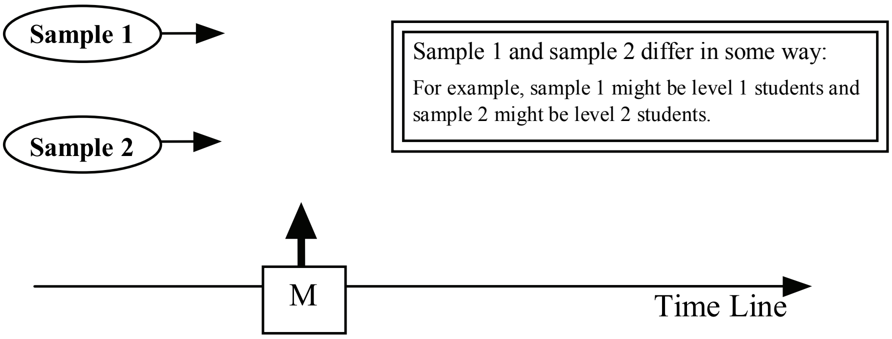


In every case, the box with an "M" inside represents a measurement of some sort (eg. a test, a questionnaire)

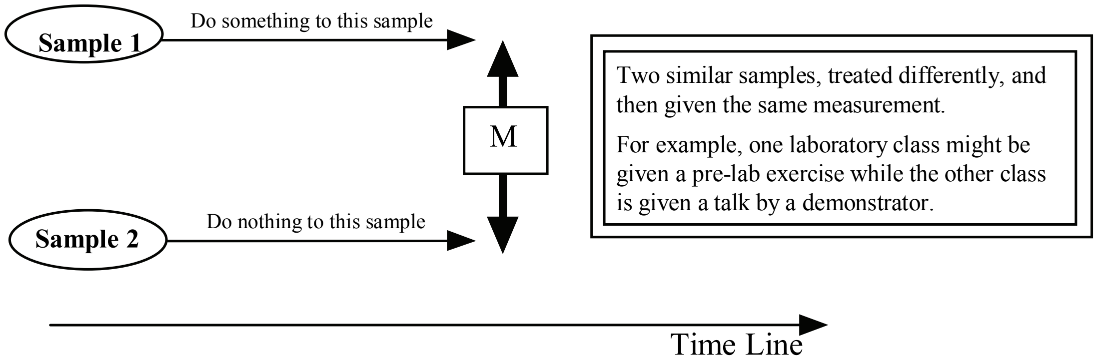
Structure 1



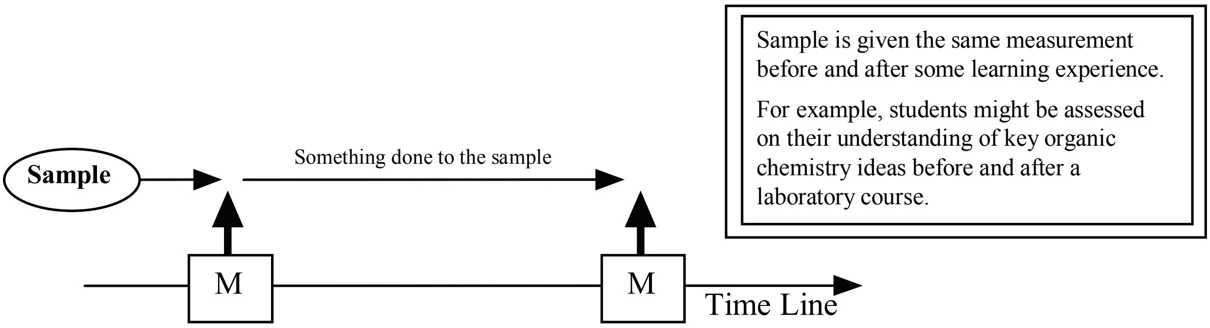
Structure 2



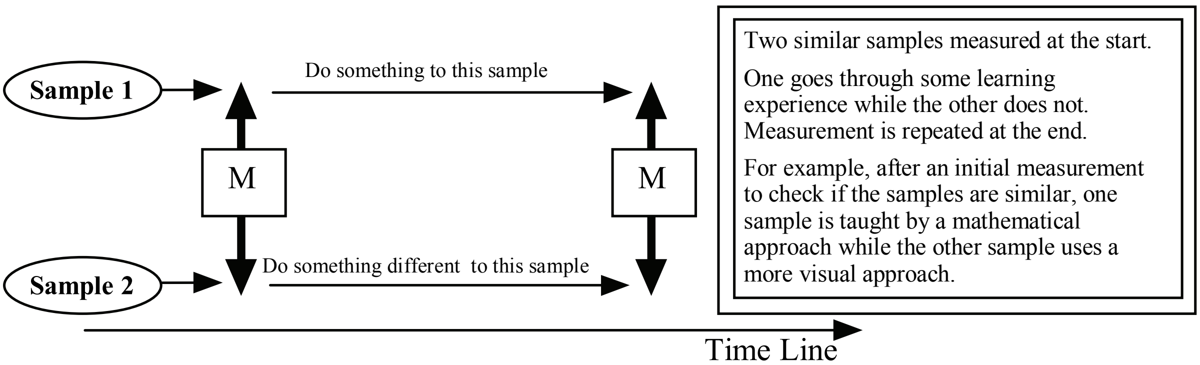
Structure 3



Structure 4



Structure 5



Using Written Tests

This section describes eleven ways of gaining insights into learning using written tests.
Most can be used in pedagogical research.
Recommendations about their individual strengths and weakness are offered.
Examples of their use are described .

Introduction

In order to assess how well learning has taken place, we set tests and examinations. In pedagogical research, tests are also used. However, the test may serve a *very different purpose*. It may have different aims and may be scored in a different way.

In university tests and examinations, we are usually seeking to assess the overall performance in a module or course. For example, we set an examination at the end of a module in astronomy to place our students in some kind of order of ability, indicating how well they have grasped our lecture course.

We may set a cut-off point and state that those who are above that point have 'passed' the course and those below have 'failed'. The cut-off point is entirely arbitrary and set by our personal judgement based on experience and the known views of others (like external examiners). No assessment can ever be absolute. Most assessments merely place our students in an order of merit based on some test we have set.

Written Tests in Pedagogical Research

In research, tests have different purposes. Here are some examples. We may wish to know:

(a) How the students have handled a specific skill;
(b) How well have our students grasped a particular concept;
(c) What specific areas of difficulty students have in tackling some theme in our teaching programme.

In other words, we are looking for *detailed insights into the student learning*. Our purpose is not to grade students, nor to award credit for some final assessment, nor to indicate who has passed (and can proceed to the next step) or who has failed (and cannot proceed onwards).

How to Devise Written tests for Pedagogical Research

There are basic guidelines for setting all tests and examinations. However, when used for pedagogical research, particular criteria may be important. These are now illustrated.

(a) ***Define as precisely as possible what you are trying to find out***. Here are some specific examples of some research questions which have been asked in previous studies.

- (i) In looking at octahedral complexes, is student difficulty dependent on the visualisation of the three dimensional structures?
- (ii) In considering transformation of coordinates in astronomy, which part of the sequence of likely mental operations is the one causing most difficulty?
- (iii) What are the key areas of misunderstanding in considering free energy as a key concept in determining reaction feasibility?
- (iv) Why is the mole a difficult concept to handle in redox reactions?
- (v) Where are the key problems in interpreting ray diagrams with lenses and mirrors?
- (vi) What has this laboratory taught my students about carbonyl reactivity?

The important thing is to specify as precisely as you can what you wish to find out. You then need to think about the evidence you would like in order to be convinced about the answer to your question. However, it is not always possible to obtain the ideal evidence: it may take too long, or it may even be that no assessment technique has yet been devised to obtain the evidence to the standard you wish.

(b) ***Select the test method which is best for the task***.
Here are some factors which might influence your choice.

With small numbers (say, <20), drawing meaningful conclusions from written methods is limited but some kind of one-to-one approach may be useful. For example, you can simply interview a selection of typical students and ask them questions relating to their learning and understanding. For more guidance on interviewing, go to pages 28-31.

Another approach with small numbers is to set a test and then talk to students on an individual or very small group (~3) basis ***while*** they are attempting the questions.

If numbers are larger (say, >50), testing on its own may be useful. You will need to devise a test which is:

- (a) Testing what you are seeking to test;
- (b) Easy to mark (your time is at a premium);
- (c) Able to be marked with consistency;
- (d) Able to be analysed statistically to gain the insights you need;
- (e) At the right level of demand for your students (if too easy or too difficult, you will learn little)
- (f) Discriminating: the students who are genuinely competent at whatever you are testing will do well while those who have real problems will do badly.

The difficult case is when numbers are between 20 and 50. Written tests will not give clear statistical evidence of patterns of learning but there is not time to interview all the candidates. In this case, it is better to use both methods but, when using one-to-one approaches, only a sample of the group will be studied.

Some Research Questions

You observe that some students seem to 'prefer' graphical methods in solving advanced flow mechanics problems while others tend to use a calculation approach: are these patterns of behaviour consistent, why do they occur, can you adjust your teaching to enable students to maximise success by adopting their preferred approach with confidence?

Some Research Questions

It is commonplace to find that students have difficulties in carrying out mole calculations when applied to quantitative analysis. What is the nature of this problem, why does it occur and what ways are possible to reduce it?

You may have found a problem with the student handling of three dimensions in stereochemistry. Is the problem related to perception of 3D, the interpretation of 2D representations into 3D, the lack of 'hands-on' opportunities with models, and is there a gender issue here?

You may wish to explore the fundamental nature of student difficulties in coping with coordinate transformations in astronomy. This fundamental skill underpins so much work and yet students find it very difficult. What are key points of difficulty? Do the difficulties lie in perception of 3D, in calculations, in conceptual understandings or what?

Kinds of Test

There are so-called objective tests, but this is a very misleading description in that they are ***only*** objective in that they can be *marked* objectively (perhaps by a machine). The most familiar will be multiple choice tests but, for research purposes, these are largely useless.

It is difficult to set good multiple choice questions and most tend to assess recall of knowledge. However, the basic weakness with multiple choice tests in the context of research is that you can never know ***why*** a student selects a wrong answer. There are numerous explanations:

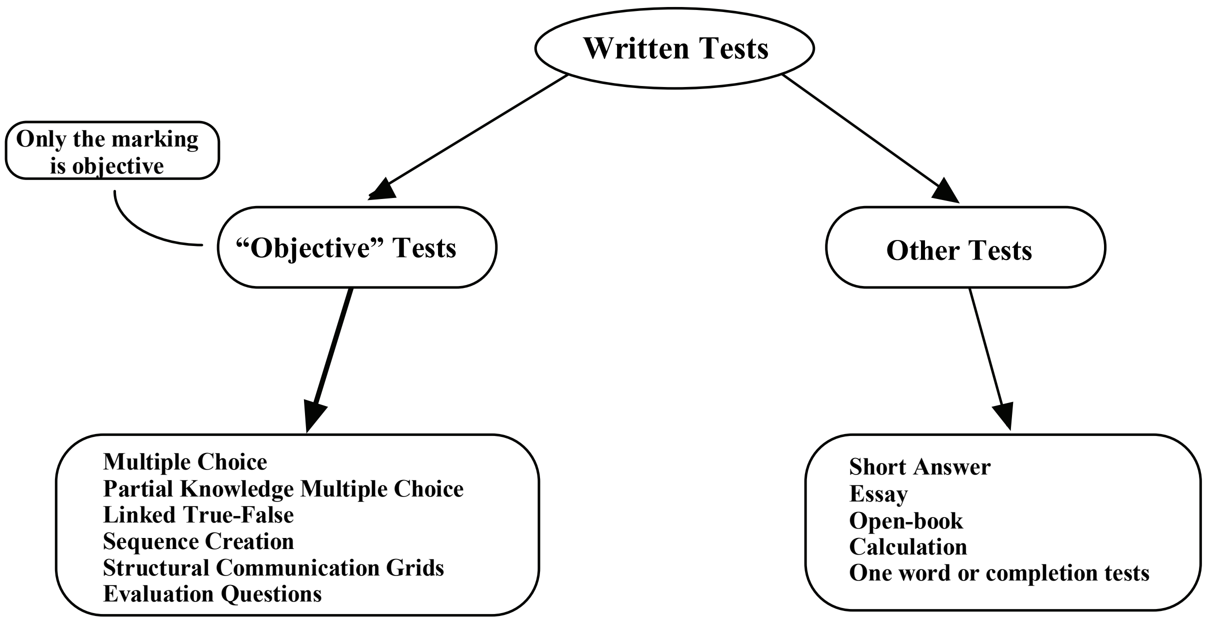
(a) They genuinely did not know the answer - but why did they choose a particular wrong answer ?
(b) They guessed and guessed wrongly;
(c) They saw something deeper in the question (not intended by you) and it led them astray;
(d) They saw something deeper in the question (not intended by you) and their understanding is deeper than measured by the question.
(e) The answer they chose was partially correct (but they still obtain zero marks).

Research has shown that 30% of correct answers chosen by students have been chosen for wrong reasons.

***Advice*:** Never use *traditional* multiple choice questions in research.

The advantage of more open-ended testing (anything from very short answers to essays) is that such methods can give you *insights* into the way the students are thinking and can reveal partial knowledge, misconceptions and areas of confusion. However, the major weakness is that marking is time consuming and it is not easy to see precisely what the patterns of students' problems may be. Objectivity of marking is not always easy.

Some Types of Tests



Each of these test types will be considered in turn.

Objective Tests

Remember, the objectivity lies only in the marking. Whether the test measures what you wish it to measure and whether the question is unambiguous and fair are both subjective decisions, determined by your professional judgement.

(a) *Partial Knowledge Multiple Choice*

This is a clever idea and awards partial marks for partial knowledge. Let us imagine we have a multiple choice question with four answers offered. Let us suppose that (B) is the correct answer. The candidate is asked to mark with a tick (✔) the correct answer and put a cross (✖) against the ***two*** answers which are definitely wrong. Consider three answers offered by three students:

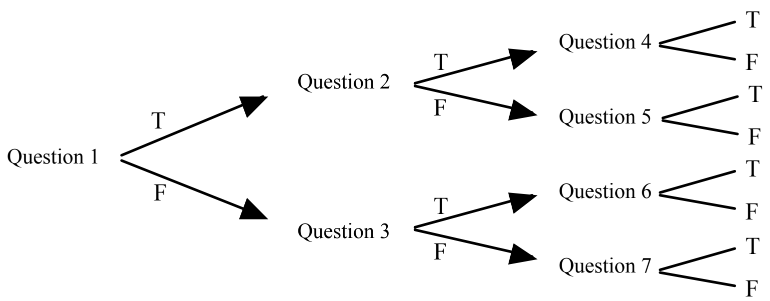
	Student 1	*Student 2*	*Student 3*
A	✖	✖	
B	✔		✖
C	✖	✔	✖
D		✖	✔

Student 1	Has ticked the correct answer and crossed two wrong answers: full marks (perhaps 1)
Student 2	Has ticked a wrong answer but has put a cross against the other two wrong answers: partial marks (perhaps 0.5)
Student 3	Has ticked a wrong answer but put a cross against the correct answer: no marks (0)

Examples can be found in: "Scoring Systems which allow for Partial Knowledge", *Journal of Chemical Education*, **55**, 1978, 717-9.

(b) *Linked True False*

In this, the choice of the student in the first question leads to one of two following questions, to each of which there are two ensuing questions. Thus the destination (the answer to the final question) of the student tells the marker precisely the route of the logic followed by the student and can give useful insights into weakness.



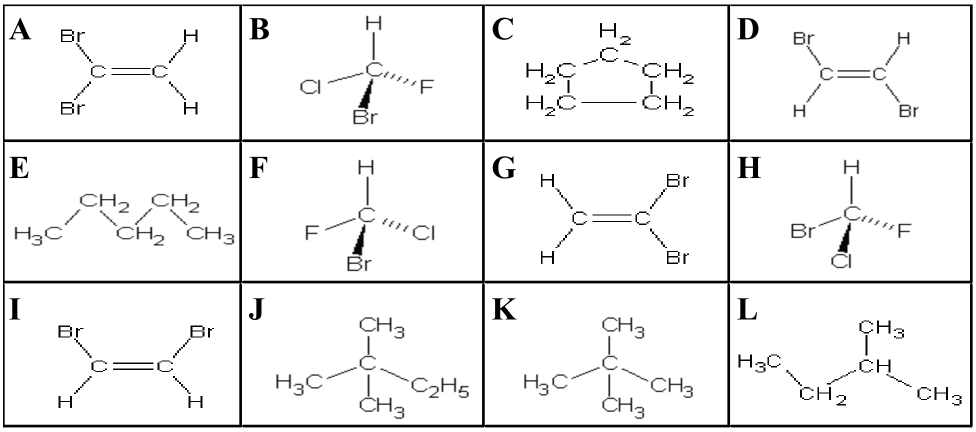
(c) *Sequence Creation*

In this, the student may be given a number of operations, or statements. The student is asked to place these in order so that the sequence of operations allows a procedure to be completed correctly. For example, this could be procedures for an experiment to be conducted meaningfully or a logical argument to be developed (as would happen in an essay).

(d) *Structural Communication Grids*

This is a very powerful and relatively easy-to-use technique, with widespread range of potential applications. Here is an example which was seeking to explore the perception and understanding of three dimensional structures in organic chemistry. This item was used at the outset of a first year course in organic chemistry which was to be using mechanisms.

Look at the boxes below and answer the questions that follow.
(Boxes may be used as many times as you wish)



Select the box(es) which contain:

(a) An isomer of the molecule shown in box A ....
(b) An isomer of the molecule shown in box D ....
(c) An isomer of the molecule L ....
(d) A molecule which is identical to the molecule shown in box F ....

There are many advantages in using Structural Communication Grids:

(a) They are much easier to set than multiple choice questions;
(b) Guessing is virtually eliminated as the student has no way in advance of knowing how many answers are expected;
(c) The correct responses offered by a student reveal something of the grasp of the fundamental concept: for example, if there are 3 correct answers but many students have missed the same one of the three, we gain clear evidence of a knowledge gap;
(d) The wrong answers offered by the students reveal something of the misunderstandings and misconceptions: for example, if many students add a particular wrong answer, this can reveal a misconception or misunderstanding;
(e) There are several ways to score such a test. Here is one:

$$\text{Mark for each question} = \frac{\text{Number of correct answers selected}}{\text{Total number of correct answers}} - \frac{\text{Number of wrong answers selected}}{\text{Total number of wrong answers}}$$

(f) Clear patterns of responses are highly informative;
(g) You can ask many questions using one grid, gaining useful insights into many aspects of some concept or area of interest.

Advice: For insights into conceptual understanding, structural communication grids are highly recommended. The grid itself has no significance other than offering a framework to 'hold' answers. The grid shown above has 12 answers and this is probably a reasonable maximum for first year undergraduates.

(e) *Evaluation Questions*

An example illustrates this. In an attempt to assess whether students have grasped the significance of some industrial chemistry related to the production of sulphuric acid, the question stated that the sulphur-producing countries of the world had formed a cartel and had agreed to raise the price of sulphur (a raw material for the process) by 20%. The question then offered a series of possible actions which could be taken to minimise the effect of this on the UK's economy. Students were asked to place these in order of relevance to meet this objective. The orders obtained offered insights into the way they understood the operation of the industry and the factors affecting production.

Summary of Some Types of Objective Tests

Some Forms of Objective Tests

Type	Advantages	Disadvantages
Multiple Choice	Easy to mark	Difficult to set; offer little useful insight
Partial Knowledge Multiple Choice	Easy to mark; give some useful insights	Fairly difficult to set
Linked True-False	Reveal unambiguous thought sequence	Limited areas of application
Sequence Creation	Reveal logic and clarity of thought processes	Limited areas of application
Evaluation Questions	Can give useful insights into understanding in a more holistic sense	Limited areas of application
Structural Communication Grids	Easy to set, mark; very versatile in application; give very useful insights	Marking needs careful thought to gain the most powerful insights

Other Types of Questions

It is possible to use questions which are not objectively marked. Six types are discussed here.

(a) *Short answer questions*

These can usefully be employed. They need to be designed in such a way that marking is reasonably assured (little room for ambiguity, judgement of marker) and conclusions can be drawn from the patterns of results. They may be useful where the information you wish can be offered in terms of short phrase or single sentence answers.

(b) *Essays*

These can offer very useful insights in pedagogical research but the time for marking is a serious drawback while the interpretation of marks may not be easy. However, if specific goals are defined clearly and the marking is determined in the light of these goals, essays can be very insightful. The power of essays can lie in the evidence they offer about logical thought, ideas sequencing and balanced overview of a theme.

(c) *Open book tests*

These can be used to explore thinking, data handling, analysis of ideas and synthesis of thought as well as important skills like creativity. The tests can take many forms but it is important to be very clear what is being explored. What such tests can do is to remove the dependence on memory (other than remembering where to find things in a textbook!), relating performance more to other thinking skills. A nice example in chemistry is to use the periodic table (with whatever information on it is relevant) and ask short questions to pick out elements or relevant information using the table.

(d) *Calculations*

In much of physical science, the ability to conduct calculations appropriately and accurately is very important. However, the danger is that students are very adept at applying such calculations algorithmically, with *little understanding* of what is going on. Interpretation of results from such tests is, therefore, potentially misleading if we are seeking to gain insights into student understanding.

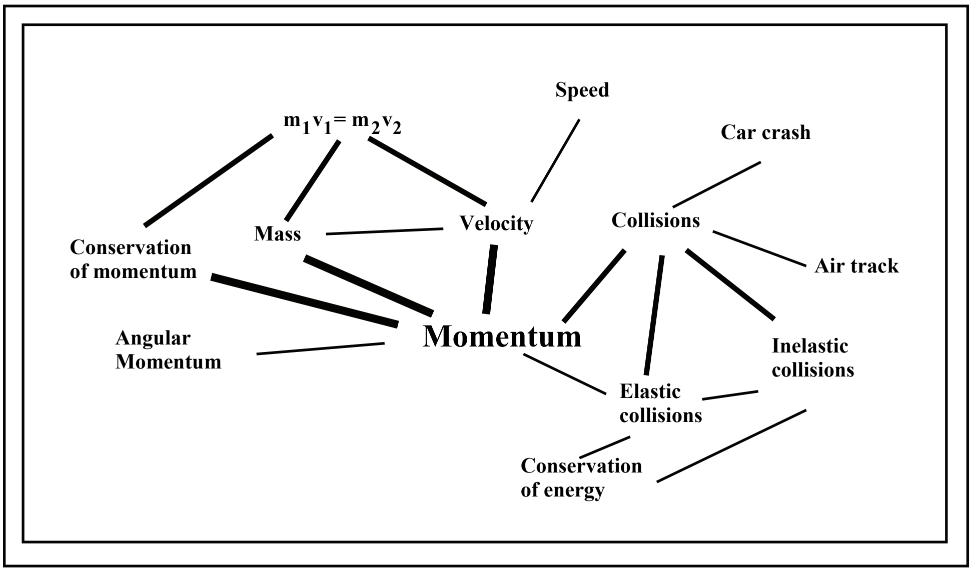
(e) *Fill-in Tests*

One word answers or answers requiring the student to fill in a word, a number or a formula can be easy to mark but may not furnish much useful evidence of the processes of understanding. They may have their place if you want insights into what has been remembered but they may not offer much evidence of understanding.

(f) *Concept Maps*

A very interesting technique which has been used to offer rich insights into the way a concept is 'mapped' in the student mind. There are two approaches described in the literature. One is to use Word Association Tests to derive a concept map (often from the whole class group). Such tests are very short, fairly easy to devise but difficult in analysis. A mind map is a pattern drawn on paper showing how ideas or concepts are linked. It is thought that the paper representation illustrates the way ideas are linked in long term memory and this can reveal how well concepts are understood.

Here is a simple example:



Let is consider a map developed from, say, 100 first year physics students. The thickness of the line shows the strength of the association. The thickest lines correspond to associations shown by, say, 50% of the students, while the thinnest of the lines might have occurred for only 5% of the student group. The map reveals the pattern of mental connections held within the class. It can inform the lecturer about connections which need strengthening or it might reveal connections which are incorrect.

An example of this approach can be found in Bahar *et al*., (1999), Investigation of Students Cognitive Structure in Elementary Genetics Through Word Association Tests, *Journal of Biological Education*, **33**(3), 134-142.

The other approach is to teach students to draw mind maps and then use these to explore how the concepts have developed in their minds. There are assertions in the literature which suggest that we can assess students on the basis of their mind maps. However, recent research casts great doubt on this[4].

Summary of Some Types of Tests

Various Other Possible Tests		
Type	**Advantages**	**Disadvantages**
Short Answer	Easy to set; can give useful insights; widely applicable	Observing patterns of responses not always easy
Essay	Easy to set; can be insightful	Difficult to mark; very difficult to see patterns
Open-book	Tests skills other than memory	Difficult to interpret results clearly
Calculation	Answers usually right or wrong	Limited in application; can have marking problems
One word or completion tests	Easy to set and mark	Reveal very little
Concept maps (Mind maps)	Very rich in insight	Interpretation requires considerable caution

4 Otis, Kevin, 2000, *Mind Mapping with Medical Students*, PhD Thesis, University of Glasgow.

Interviews

This section describes the potential of interviews.
It outlines how to plan and conduct interviews.
Examples of their use are described.
Ways to analyse data obtained from interviews are presented.

We can learn a large amount by ***talking*** to our students about their learning. Interviewing is a powerful research tool. However, there are several possible pitfalls:

(a) Interviews take time both for students and for staff: a useful interview make take anything from 15-30 minutes. Interviewing, say, 20 students, then takes anything from 4 to 8 hours of staff time.
(b) It is difficult to translate evidence from interviews into a neat summary: summarising interviews from, say, 20 students into a meaningful whole is a time-consuming task and, because each is different in terms of the way the students use language and ideas, it is not always easy to be sure of common ideas.
(c) It is possible to allow interviews merely to confirm our preconceived ideas: Even with great effort, it is difficult to avoid the fact that, if the member of staff is the interviewer, our ideas and thoughts may influence the way the interview is conducted and the way results are interpreted.
(d) Finally, if we are perceived as some kind of 'authority figure' in our departments, then it is possible for students to respond in ways which they perceive as putting themselves in a good light.

How do we get round these potential pitfalls ?

It is important to be very clear ***why*** we wish to interview. Many of the practicalities resolve themselves after that.

Here are two examples from some research which illustrate some ways forward.

Example 1
In a major study of student attitudes towards physics in the first two years of university study, a questionnaire had been used. A small number of interviews were then set up for three purposes:

(i) To check if the way the students had responded in the questionnaires was, in fact, consistent with what they really thought;
(ii) To explore one idea further: students had indicated the kind of physics they wished more of during their studies and it was thought important to amplify and clarify what they were saying;
(iii) To explore further some ideas about laboratory work. This was an area where one year group had indicated considerable dissatisfaction and the researcher had some new ideas to be discussed.

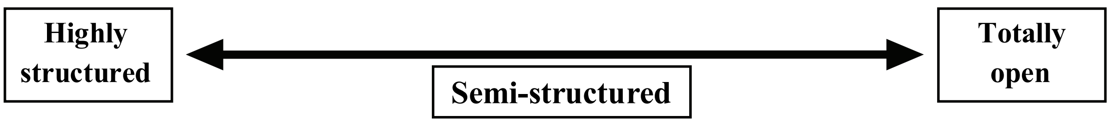
After a few minutes of general questioning to establish a relaxed atmosphere, 10 minutes was allocated to aims (i) and (ii), leaving 15 minutes to explore (iii). The interviewer used a semi-structured approach, with a set of fixed questions merely as a starter framework. Much of the interview was open, with the students encouraged to talk freely.

Because the questionnaire data had been gained from a large number of students, the interview was being used to check that the questionnaire was revealing what was intended and to explore two areas where the questionnaire had indicated interesting patterns of responses. The interviews were being used to amplify the questionnaire data. Only a small sample was needed for these purposes (in fact 13 were interviewed), 30 minutes being allowed for each.

Example 2
In another study looking at laboratory work in a first year chemistry course, the organisation of the laboratory was studied, manuals were scrutinised and a questionnaire was applied to hundreds of students. The researcher then spent time in the laboratories, working alongside a demonstrator. After a week or two, now accepted as a part of the 'laboratory furniture', he was able to talk informally to about 60 students during the laboratory time over many weeks and, in a casual but systematic way, was able to build up a check list of student attitudes towards specific features which were of interest to him.

How to Interview

Interviews can be ***highly structured***, with all the questions decided beforehand. This is more like a verbal questionnaire but has the advantage over a questionnaire of allowing some kind of check for misunderstanding and misinterpretation. If you plan to use several interviewers, this can only be done if there is high measure of structure, if there is a planning session with all interviewers in advance and if there is complete agreement on how the interviews will be conducted. Using more than one interviewer reduces personal biases but opens the door to inconsistency and ensuing difficulties in data analysis.



They can be ***totally open***, built around a question like: "How did you find the physics laboratories this year?" or "Tell me what you find the best way to learn organic chemistry?" Some preliminary questions may be needed to develop the levels of confidence and trust to enable the student to talk freely and openly. This type of interview is unpredictable and can be long. However, the insights gained can be very rich with an experienced interviewer who can encourage the students to talk freely.

Given the right atmosphere of confidence, this kind of interview allows the student to shape the agenda and can be realistic and revealing, if very difficult to summarise. Indeed, in appointment interviews, an opening question like, "Tell me a bit about yourself...what are the major activities you have been involved in the past year or so...." can be very helpful in that the candidate reveals the priorities they perceive in their present work and, being their own work, they can be confident and relaxed, and maybe even enthusiastic

Interviews can be semi-structured, with a series of fairly well defined questions but with much time left for open discussion, determined by the way the student reacts or wishes to go. This allows some measure of student freedom but, if conversation dries up, the interviewer can fall back on the next question.

In building up a set of questions to be posed, there are a number of general principles:

(a)	**Relate your questions tightly to your specified aims;**
(b)	**Keep language straightforward and be prepared to ask questions in two or more ways;**
(c)	**Have 'reserve' questions for the shy interviewee;**
(d)	**Seek a second opinion from a colleague on your questions;**
(e)	**If possible, try the questions out on a suitable 'guinea-pig'.**

Practical Issues

One of the key aspects of interviewing is the interviewer-interviewee relationship. If the interviewee suspects some kind of 'agenda', responses may not be completely honest!

This leads to the question about who conducts the interview. *A socially competent research student may be a far more useful interviewer than a staff member!* In most universities, there are teaching support departments who can provide an experienced interviewer. Another source of interviewer might lie in a Psychology Department where honours students have to conduct honours projects. Frequently, there may be students with an interest in learning who can be excellent interviewers. The use of such interviewers may reduce many problems.

Interviews need not necessarily be one-to-one. If the group is too large (perhaps 4 or more), then it is likely that there will be 'passengers' in the group who will contribute very little. Interviewing one-to-one requires some interviewing skill in that a shy or hesitant student may need coaxing and encouragement. With a group of two or three, students are often more relaxed and can support each other. However, it is not always easy to avoid the views of one student influencing the others and perhaps dominating the interview direction. There is no simple best way. It depends on your aims, the time available and the nature of student social dynamics.

The way students react will also be influenced by the location of the interview. Informality in familiar surroundings is strongly advocated. One of the best ways observed is when research students carry out the interviews, with cans of juice, cups of coffee and nibbles on hand. Relaxed honesty comes quickly! It is worth avoiding the formality of an interviewer and an interviewee eyeing each other over the table.

It is important to be up front with the students and to explain the purpose of the interview. The interview (in a research setting) is a voluntary affair. The student needs to know what is happening, why the interview is taking place and what use will be made of the outcomes which will emerge from the interviews. Anonymity is important. Some departments have specifically asked that it be made clear to students that the survey of student views is being conducted by someone *outside* the department and no view will be able to be related to a specific student. This may be helpful although, where staff-student relationships are confident and open, this may not be necessary.

Inevitably, the interviewer will make notes during the interview, unless blessed with a photographic memory. It is important to ensure that note taking is as unobtrusive as possible and does not hinder the flow of the interview.

Some people conduct interviews with a tape-recorder running throughout. This needs some thought in advance in that analysing a tape of an interview may take several hours for each hour of taping. However, such tapes can be useful to amplify any notes made by the interviewer. It may be useful only to refer to those tapes where a particularly interesting discussion was observed. If a tape recorder is to be used, it is essential that it is unobtrusive and that the microphone is sensitive (teaching support services or media support services can usually help here). Time has to be allowed at the outset for the interviewees to ignore the taping and some interviewees may not wish to be recorded and this must be respected.

Planning Interviews

(a) Define your purpose precisely.
(b) Decide on level of structuring and be willing to modify as necessary; prepare list of questions
(c) Check on time available for interviewing and data analysis;
(d) In the light of aims, decide in general terms how you propose to analyse data;
(e) Plan how many will be interviewed, how the sample will be obtained and whether it will one-to-one or in small groups;
(f) Decide who will be the interviewer;
(g) Decide where and when and the context (eg. over coffee).

Analysing the Outcomes

If the interview is highly structured, then data analysis can be simpler. For example, we can simply record the proportion of students who like laboratories, used a given textbook regularly, re-wrote their lecture notes after lectures and so on. However, most interviews are, in some measure, not so structured and students can respond in widely different ways using widely different language. Indeed, in open interviews, the student may even determine the agenda.

It all comes back to the original aims. Here is an example which illustrates this:

> In a series of student interviews, the aim was to explore how students saw the process of problem-based learning which was used in a particular course. The aim was exploratory. There was no decision beforehand about possible student reactions. The interview was highly unstructured and fairly long. Interviews were recorded and transcribed and each student was treated as a case study. Such work can be highly revealing although very time-consuming.

This is completely different to an analysis of responses in categories as in a highly structured interview. In many interviews, there is partial structuring. Different parts of the interview will be treated in different ways. Responses to some questions may simply be totalled while responses to others may vary widely. A summary of the interviews might look like:

> "All 10 students were clearly very much in favour of the new laboratory exercises. 6 indicated that they like the way they related to real-life issues in optics while 3 felt that it was important that the marks awarded should reflect the effort required. 8 want the same approach to be applied to all the experiments, 4 like the opportunity to be creative although 2 others express their slight reservation about how they might cope."

Because the sample was only 10, care must be taken in generalising the results and statistical analysis is completely inappropriate.

Attitudes

This section describes the nature of attitudes relating to the physical sciences.
It outlines the main ways attitudes can be measured.
It offers a guide in designing a questionnaire.
Examples of their use are described.

What are Attitudes ?

Attitudes express our ***evaluation*** of something or someone. They may be based on our knowledge, our feelings and our behaviour and they may influence future behaviour. In the context of studies in the sciences, attitudes are evaluations which may influence thinking and behaviour.

An attitude must have a ***target***. We have an attitude directed towards something or someone. Attitudes are highly complex and can affect learning extensively.

Some Examples

(a) There have been numerous studies looking at attitudes towards physics and the study of physics. This is because it is known that a student's attitudes towards physics (and all the aspects of that study) is the strongest factor which will influence whether the student will later choose to continue study with physics. By analysing the way student attitudes towards physics develop, it is hoped that more might be enabled to choose physics as an area of study.

(b) There are many kinds of fairly specific attitudes. Students will develop attitudes towards laboratory work, towards certain studies of teaching and learning, towards their teachers, their demonstrators, specific topics: the list is endless.

(c) During a course of study, students will also develop attitudes towards themes and topics which they study. For example, as a chemistry student learns more about chemical industry through their course, they will develop attitudes towards aspects of the work of chemical industry. As students develop understandings of space physics and astronomy, they may well develop attitudes towards aspects of man's involvement in space. Studying topics which involve work-related issues (for example, pollution, nuclear industry, medical physics, and drug development) will inevitably allow students to develop attitudes towards these and related themes.

(d) Of course, students will have attitudes towards study and work. For many, this will mean minimising the demands placed on them. However, there is considerable research which has explored the way students see the role of their lecturers, themselves, the assessment and the actual nature of the knowledge to be learned. Such attitudes are extremely important in developing graduates who will not only be successful in their studies at university but also be able to continue learning throughout the rest of their lives.

(e) We sometimes refer to the scientific attitude. This is not so much an attitude as a way of working and will be discussed later.

Why are Attitudes Important ?

The first thing to recognise is that students will develop their own attitudes - it's going to happen anyway! If we ignore attitudes in our thinking about teaching and learning, that will not stop the students developing attitudes. If we think that our task is to communicate astronomy, chemistry and physics ideas and nothing more, that still will not stop attitudes developing.

Attitudes are important to us because they cannot be neatly separated from study. It is a relatively quick series of steps for a student with difficulty in a topic to move from that to a belief that they cannot succeed in that topic, that it is beyond them totally and they, therefore, will no longer attempt to learn in that area. A bad experience has led to a perception which has led to an evaluation and further learning is effectively blocked.

In general, attitudes in life allow us to:

(a) *Make sense of ourselves;*
(b) *Make sense of the world around us;*
(c) *Make sense of relationships.*

Of course, we want our students to make intellectual sense of the world around them - that is the very nature of the subject matter of the physical sciences (and other sciences). Of course, it helps them to make contributions to the understanding of the world if they can also make sense of themselves and others.

Four Targets

There are four broad areas where we might wish to explore attitudes in relation to students

(a) *Attitudes towards subjects being studied;*
(b) *Attitudes towards study itself;*
(c) *Attitudes towards the implications arising from themes being studied;*
(d) *The so-called scientific attitude.*

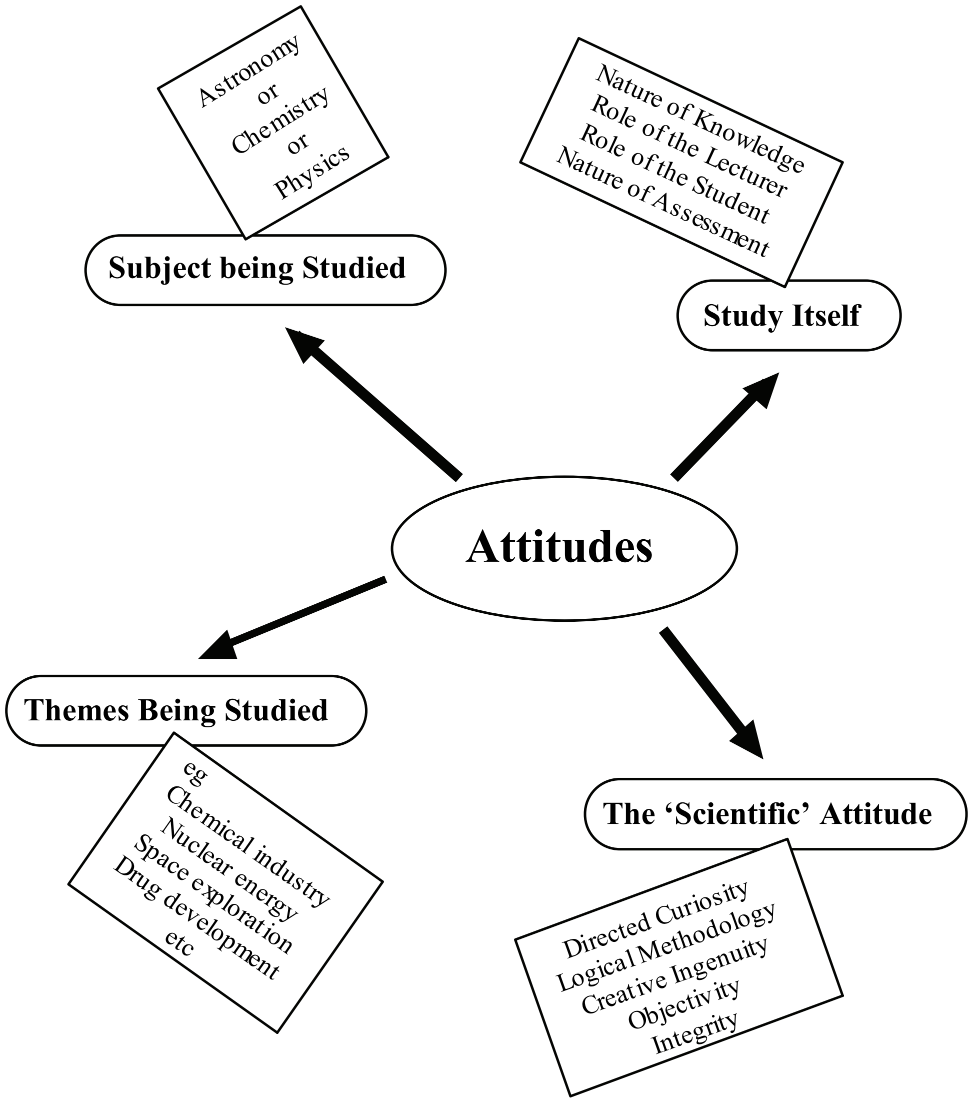
Considerable work has been undertaken looking at attitudes towards subjects like chemistry and physics. Most of this has been in the USA where there is continual concern about the popularity of these subjects. There has also been considerable interest in the ways students study and their attitudes to study. Some workers have focussed on deep and surface learning, with much of the research work being qualitative in nature. A very useful analysis was carried out by William Perry[5] in Harvard University and this has led to a useful framework for analysing students' attitudes to work under four headings:

(a) *Student's perceptions of the nature of knowledge;*
(b) *Student's perceptions of the role of the lecturer in their learning;*
(c) *Student's perceptions of their own role in learning;*
(d) *Student's perceptions of the nature and role of assessment.*

---

5 Perry, W.G., (1999), Forms of Ethical and Intellectual Development in the College Years: a scheme, San Francisco, USA: Jossey-Bass Publishers.

A Summary



Astronomy
or
Chemistry
or
Physics
Subject being Studied
Nature of Knowledge
Role of the Lecturer
Role of the Student
Nature of Assessment
Study Itself
Attitudes
Themes Being Studied
eg
Chemical industry
Nuclear energy
Space exploration
Drug development
etc
The 'Scientific' Attitude
Directed Curiosity
Logical Methodology
Creative Ingenuity
Objectivity
Integrity

How can we Measure Attitudes?

Although many felt that attitudes were ascertainable, for many centuries it was thought that attitudes could ***not*** be measured. It was around 1929 when the first serious attempt was made, to be followed by the work of a researcher called Likert who has given his name to a technique which is widely used today.

Ideally, if attitudes lead to behaviour, then we aim to measure behaviour and then deduce what the attitude might be. A simple example is to look at the numbers who choose to study physical sciences. In schools in most countries, physics is not seen as popular and this negative attitude for many is easily observed in their behaviour: they choose ***not*** to continue with physics studies and, in England and Wales, A Level entries have fallen.

From the data on A Level entries over many years, it is possible to make some deductions about school pupil attitudes to physics. Of course, if they have opted out at that stage, it is highly unlikely that they will reappear in physics courses at university. This example has been chosen because, in two countries at least (The Netherlands and Scotland), this pattern of attitudes towards physics is ***not*** observed. Numbers choosing to stay with Physics at school has grown over many decades. Nonetheless, in Scotland, the numbers who choose to study physics at university has ***not*** grown. Here, it has been shown that the school pupils have an interesting attitude towards physics. They see it as a vital stepping stone to studies in related disciplines like engineering, computing and electronics as well as the medical sciences. Sadly, their perception of opportunities in physics itself is less positive. All this illustrates why attitudes are so important.

To assess our students' abilities in, say, chemistry, we either score a written test or we subject them to an oral examination. When we want to find out if a candidate is suitable for a post we are seeking to fill, we assess them on their written CV and application form and then we interview them.

In essence, the same two main ways are used to gain insights into student attitudes:

(a) We talk to them - some kind of organised interviewing;
(b) We ask them to respond to some kind of written 'test'.

Written Tests

These are often called questionnaires and there is a common view that such questionnaires are highly unreliable and of limited value. The evidence shows that this need ***not*** be the case. A well-constructed questionnaire can provide extremely accurate insights into how students think and the way they evaluate situations and experiences.

On the following pages are some examples of approaches which have been found useful in pedagogical research in recent years. They are parts of questionnaires which have been used with large numbers of students.

(a) ***Semantic Differential*** (based on the work of C.E. Osgood[6])

What are your opinions about your laboratory experiences in chemistry ?
Tick ONE box on each line.

Useful	☐	☐	☐	☐	☐	☐	Useless
Not helpful	☐	☐	☐	☐	☐	☐	Helpful
Understandable	☐	☐	☐	☐	☐	☐	Not understandable
Satisfying	☐	☐	☐	☐	☐	☐	Not satisfying
Boring	☐	☐	☐	☐	☐	☐	Interesting
Well organised	☐	☐	☐	☐	☐	☐	Not well organised
The best part of chemistry	☐	☐	☐	☐	☐	☐	The worst part of chemistry
Not enjoyable	☐	☐	☐	☐	☐	☐	Enjoyable

What are your opinions about University Physics?
Place a tick in one box between each phrase to show your opinions.

I feel I am coping well	☐	☐	☐	☐	☐	☐	I feel I am not coping well
I am not enjoying the subject	☐	☐	☐	☐	☐	☐	I am enjoying the subject
I have found the subject easy	☐	☐	☐	☐	☐	☐	I found the subject hard
I am growing intellectually	☐	☐	☐	☐	☐	☐	I am not growing intellectually
I am not obtaining new skills	☐	☐	☐	☐	☐	☐	I am obtaining new skills
I am enjoying practical work	☐	☐	☐	☐	☐	☐	I am not enjoying practical work
I am getting worse at the subject	☐	☐	☐	☐	☐	☐	I am getting better at the subject
It is definitely 'my' subject	☐	☐	☐	☐	☐	☐	I am wasting my time in this subject

(b) ***Likert*** (based on the work of R Likert[7])

Think about your experiences in laboratory work in chemistry.
Tick the box which best reflects your opinion.

		Strongly Agree	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
(a)	I believe that the laboratory is a vital part in learning chemistry	☐	☐	☐	☐	☐
(b)	I prefer to have written instructions for experiments	☐	☐	☐	☐	☐
(c)	I was unsure about what was expected of me in writing up my experiment	☐	☐	☐	☐	☐
(d)	Laboratory work helps my understanding of chemistry topics	☐	☐	☐	☐	☐
(e)	Discussions in the laboratory enhance my understanding	☐	☐	☐	☐	☐
(f)	I only understood the experiment when I started to write about it afterwards	☐	☐	☐	☐	☐
(g)	I had few opportunities to plan my experiments	☐	☐	☐	☐	☐
(h)	I felt confident in carrying out the experiments in chemistry.	☐	☐	☐	☐	☐
(i)	I found writing up about experiments pointless	☐	☐	☐	☐	☐
(j)	The experimental procedure was clearly explained in the instructions given	☐	☐	☐	☐	☐
(k)	I was so confused in the laboratory that I ended up following the instructions without understanding what I was doing	☐	☐	☐	☐	☐

6 Osgood, C.E., Suci, C.J., and Tannenbaum, P.H., (1957), The Measurement of Meaning, Urbana, Ill: University of Illinois Press.
7 Likert, R., (1932) A Technique for the Measurement of Attitudes, Archives of Psychology, 140, 5-53.

(c) An example of part of a questionnaire based on the work of Perry[8].

Students are asked to tick one box on each line.

In order to pass my courses, I need to study just what the lecturer tells me							I do not have to rely totally on the lecturer. Part of my learning is to work things out myself
I do not believe in just accepting what the lecturer says without question. Success involves thinking for myself.							I cannot be wrong if I accept what the lecturer says. If I question anything, I might end up failing
I believe it is the job of the lecturer to supply me with all the knowledge I need							The duty of the lecturer is not to teach me everything, but to help me to think for myself.
All one has to do in science is to memorise things.							Understanding science is the key part of science study.
I do not believe that all scientific knowledge represents the 'absolute Truth'.							We cannot call anything scientific knowledge if it is not absolutely true.
I do not like short questions as they do not give me the chance to explain what I know and understand.							I prefer to learn the facts and then be tested on them in short questions
In exams I prefer questions which are based on what the lecturer taught.							In exams, I like questions that give me the scope to go beyond what is taught and show my ability to think

	Strongly agree	**Agree**	**Uncertain**	**Disagree**	**Strongly disagree**
Sometimes I find I learn more about a subject by discussing it with other students than I do by sitting and revising at home					
There is not any point in class teaching, which includes things which will not be in the exam					
If I have the choice of written comments or a specific mark at the end of a piece of science coursework, I would choose the comments					
It is a waste of time to work on problems which have no possibility of producing a clear-cut, unambiguous answer					
A good thing about learning science is the fact that everything is so clear-cut: either right or wrong					
I like exams which give me an opportunity to show I have ideas of my own					

The advantages of the Osgood method are its ease of construction, the speed at which it can be answered, and the fact that both ends of the scale (in the above examples, a six point scale) are defined. However, there are limitations to its usefulness without it becoming too wordy.

The advantage of the Likert method is its versatility but only one end of the scale (in the above examples, a five point scale) is defined.

Both methods are recommended. Six and five point scales are appropriate.

---

8 Perry, W.G., (1999), Forms of Ethical and Intellectual Development in the College Years: a scheme, San Francisco, USA: Jossey-Bass Publishers.

(d) ***Other written forms***

Here are several reasons why laboratory work is part of most chemistry courses
Place a tick against the THREE reasons which YOU think are the most important.

Chemistry is a practical subject ☐
Experiments illustrate theory for me ☐
New discoveries are made by means of experiments ☐
Experiments allow me to find out about how materials behave ☐
Experiments teach me chemistry ☐
Experimental skills can be gained in the laboratory ☐
Experiments assist me to plan and organise ☐
Experimental work allows me to think about chemistry ☐
Experimental work makes chemistry more enjoyable for me ☐
Laboratory work allows me to test out ideas ☐

This question was looking to see how students saw the purposes of laboratory work.
The frequencies of ticks in the boxes gave a pattern which allowed the researcher to determine the order of importance of the various reasons for laboratory work in chemistry.

Which factor(s) influenced your choice of planned honours subject(s)?
Tick as many as you wish

☐ Enjoyment of subject
☐ Good grades at school in subject
☐ Your teacher at school
☐ Your parents
☐ Information from mass media
☐ Friends
☐ Likely career opportunities
☐ Demonstrations, exhibitions, festivals
☐ Any other factors (please list below)
....

This question was seeking to explore which factors had influenced students towards physics and related subjects. The appendix outlines the outcomes from such a question and the way the data were handled.

(e) ***Interviews***

Interviews can be a powerful tool to gain insights into student attitudes. They can also be used with samples of students after a questionnaire has been used to gain further insights into the results which have been obtained from the questionnaire. A full discussion of interviews, their use, their planning and their place in pedagogical research is given on pages 28-31.

Interpreting our Observations

Imagine we have set a very simple question to explore student attitudes towards laboratories in physics. We have given this question (in a longer questionnaire) to both first year and second year students. Let us look at six questions from this questionnaire.

Tick the box which best reflects your opinion.	*Strongly Agree*	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
(a) The experimental procedure was clearly explained in the instructions given	☐	☐	☐	☐	☐
(b) Laboratory work helps my understanding of physics topics	☐	☐	☐	☐	☐
(c) I had few opportunities to plan my experiments.	☐	☐	☐	☐	☐
(d) I felt confident in carrying out the experiments in physics	☐	☐	☐	☐	☐
(e) I was unsure about what was expected of me in writing up my experiment	☐	☐	☐	☐	☐
(f) I was so confused in the laboratory that I ended up following the instructions without understanding what I was doing.	☐	☐	☐	☐	☐

Suppose we obtain completed questionnaires from:

145 level 1 students (105 men and 40 women)
107 level 2 students (78 men, 29 women)

Let us look at the data which might have been obtained for the first three questions.
This is shown as percentages.

Level 1 145 students

	%	%	%	%	%
(a) The experimental procedure was clearly explained in the instructions given	21	35	24	9	11
(b) Laboratory work helps my understanding of physics topics	6	8	31	35	20
(c) I had few opportunities to plan my experiments.	45	32	14	8	1

Level 2 107 students

	%	%	%	%	%
(a) The experimental procedure was clearly explained in the instructions given	15	30	28	18	9
(b) Laboratory work helps my understanding of physics topics	8	14	34	31	13
(c) I had few opportunities to plan my experiments	34	29	23	8	6

It looks like level 2 students hold different attitudes when compared to the level 1 students but this could have happened by chance.

The way to check if this could have happened by chance is to use a statistic called chi-square (χ^2). The method is described in appendix 2 and most statistical computer packages will calculate the result quickly. It turns out that here the value for χ^2 for question (a) is 13.3 and tables show that this value indicates that the level 2 student responses to this question are unlikely to have arisen by chance when compared to the level 1 students. Question (b) give a χ^2 value of 8.9 which suggests that level 2 results are very likely to be not different from level 1 while, in question (c), the χ^2 value is 12.4 which indicates that level 2 is likely to be different from level 1 in responding to the question.

It is also possible to compare the men and the women for each year group.

> ***Warning*:** **χ^2 is not quite as straightforward as it might appear!**
> **If you want more details of some of the hazards in using this statistic, see appendix 2**

Attitudes and Attitude Scales

Attitude scale techniques can be used. Because the literature abounds in such work, this approach will be described briefly and arguments offered to show that this is probably ***not*** very helpful in pedagogical research in the physical sciences.

Here is a simplified imaginary example.

Suppose we want to look at attitudes towards chemistry with first year students. We develop about 20 statements which relate to the study of chemistry and ask students to respond to these on a five point Likert type scale. Here is one such statement:

	Strongly Agree	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
Chemistry laboratory work is a pointless waste of time	☐	☐	☐	☐	☐

We award:

We award:	5 points	for a 'strongly disagree' (this is the most positive attitude)
	4 points	for an 'disagree'
	3 points	for neutral
	2 points	for 'agree'
	1 point	for 'strongly agree'

Stop:

This assumes that our scale is linear and the five points are equidistant

This implies that a 'disagree' is worth twice an 'agree' - a meaningless statement.

This is shown:

	Strongly Agree	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*
Chemistry laboratory work is a pointless waste of time	☐	☐	☐	☐	☐
	1	2	3	4	5

All the questions are marked in a similar way. The marks for each student are added up to give a total score. The highest scores is said to indicate the most positive attitude.

Quite apart from the dubious ways points are allocated, the technique is highly suspect. For example, two questions might both perform well in that those who score highest also score highest in the total test. One question might be asking about chemistry laboratories while the other question might be asking about the importance of the chemical industry. 'Adding' these two scores gives a meaningless total because the two questions, while highly correlated, are completely different in subject matter. Adding up a set of such scores may give a number but that number may be fairly meaningless and all the interesting patterns of responses for individual questions are lost.

Advice: It is strongly recommended that this approach is NOT used. It abuses the nature of number, makes all kinds of unwarranted assumptions and applies statistics based on certain features of data which are frequently untrue.

An Illustration

Suppose we wish to gain some insight into our health and fitness. We book into a clinic where all kinds of measurements are made: blood pressure, temperature, heart rate, heart rate when exercising, height, weight, cholesterol level, subcutaneous fat levels, and so on. We come away with a printout of all these measurements and some recommendations from the nurse ringing in our ears.

It would be folly to add up the data on our print out!

We cannot add weight to blood pressure to heart rate and so on. The 'sum' would be utterly meaningless

It is strongly recommended that, in analysing a questionnaire, each question be taken on its own and interpreted separately.

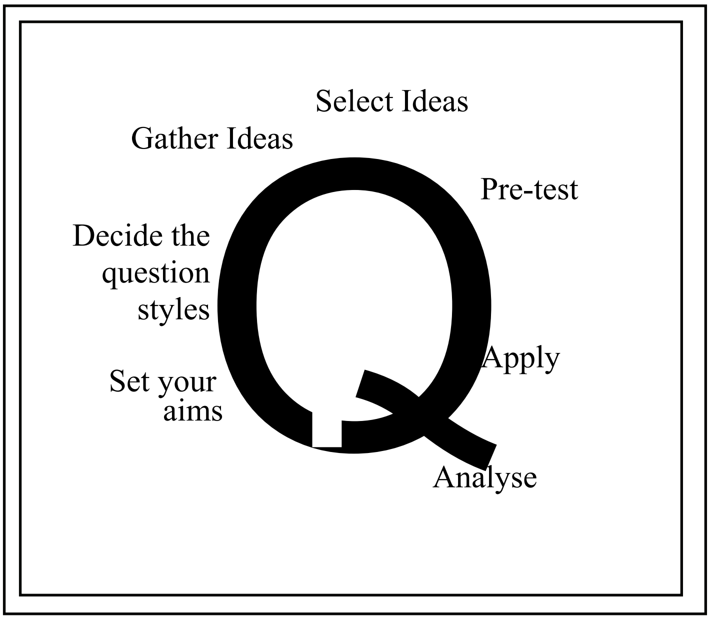
General Guidance

Developing a good questionnaire is an art and it takes a good deal of practice to achieve good results. Here is a set of procedures which can help.

- **(a) Write down as precisely as possible what you are trying to find out;**
- **(b) Decide what types of questions would be helpful;**
- **(c) Be creative and write down as many ideas for questions as you can;**
- **(d) Select what seem the most appropriate from your list - keep more than you need;**
- **(e) Keep the English simple and straightforward, avoid double negatives, keep negatives to a reasonable number, look for ambiguities, watch for double questions;**
- **(e) Find a critical friend to comment on your suggested questions;**
- **(f) Pick the best, most appropriate and relevant questions, thinking of time available;**
- **(g) Layout is everything !!**
- **(h) Try your questionnaire out on a small sample of students (eg. a tutorial group) - ask for comments, criticisms. Check time required.**
- **(i) Make modifications and only then apply to larger group;**
- **(j) Analyse each question on its own (see section on statistics, especially the use of chi-square).**

Remember, even with all that, you will almost certainly find one or two questions which are flawed in some way. Perfect questionnaires are a virtual impossibility.

The procedure can be summarised as



Handling Data Obtained

This section looks at the kind of data which pedagogical research is likely to produce.

It looks at the kind of statistics which may be useful in interpreting data.

It does not offer details of specific statistical techniques.

The Purpose of Statistics

The purpose of statistics may be to

(a) Describe what has been observed in a form which makes more sense to the reader; (*Descriptive Statistics*)

(b) Explore what data might mean (*Inferential Statistics*)

Types of Data Obtained

In pedagogical research, you will obtain information related to the processes of teaching and learning. The information can be in many forms:

(1) ***Ratio Data***: You may have information where there are scores which are on a scale like a temperature scale, using Kelvins. There is a zero and the interval between the values are equal: 400K is twice 200K. For example, you may be considering data like age or the number of examination passes at "A" Level or Higher Grade.

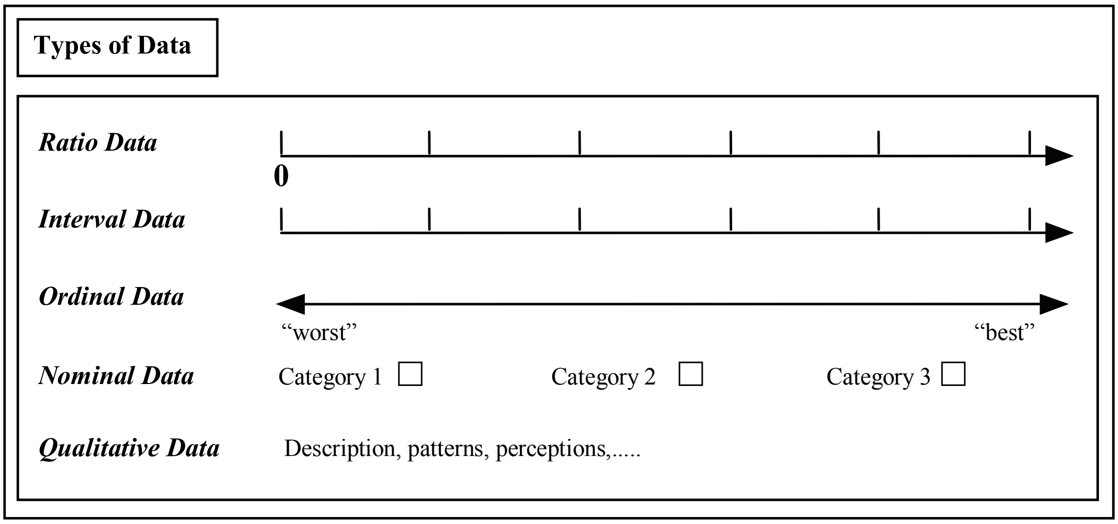
(2) ***Interval Data***: This is perhaps more common and can be used to describe the typical marks you obtain in a test or examination. There is an attempt to make the intervals between marks equal so that the gap between 40 and 50 is similar to the gap between 70 and 80. However, there is no certainty that zero on your mark scale means that the candidate knows nothing!!

(3) ***Ordinal Data***: This can also be found frequently in test and examination marks. A candidate who is awarded an "A" is better than one awarded a "B" and so on but the gaps between awards are not fixed clearly. Similar kinds of information can be obtained when measuring attitudes where all we can say is that the attitude is more or less positive but we cannot say by how much with any certainty.

(4) ***Nominal Data***: This kind of information is where we are noting categories, like male and female, mature student, student straight from school and so on.

(5) ***Qualitative Data***: This is where the information you have obtained cannot easily be converted into any kind of number system. It includes useful information like descriptions of student behaviour during tutorial sessions, behaviour in laboratories, how students see their role in learning and so on.

Types of Data



Two Important Principles

(a) *You can only use the statistics (and mathematics) appropriate to the kind of data you have obtained.*
(b) *All data are prone to error.*

If we look at the five kinds of information which you might obtain, mathematical manipulation is more or less completely inappropriate to category 5 altogether. Only very limited mathematics can be applied to category 4. For example, if you have test marks (as in category 2), then it may be useful to work out an average (mean). However, you cannot work out an average for data which is ordinal or nominal (categories 3 or 4): there is no average gender, no average for maturity, no average (mathematically) for an attitude!

One of the greatest sources of error is to use the wrong statistics for a given set of data. A typical example is when a student group is awarded As, Bs, Cs, Ds, Es in some measurement. This could be an end of course grade, an essay mark or a measure on some attitude. The temptation is try to find an average and to call an A=5, B=4, C=3 and so on. However, this assumes that the five categories are equally spaced. If that assumption cannot be sustained, then the calculation of an average is meaningless. Also, an average mark may mask the importance of distribution. A 'bell' distribution will give the same average as any other symmetrical distribution.

The second principle is also important. In the physical sciences, when we measure anything, we take into account the possibility of error in our measurement. We may frequently ask our students to estimate the extent of that error. The same principle applies to our pedagogical measurement. Our measuring instruments may not be oscilloscopes and pH meters. They may be all kinds of tests and examinations and they will contain inherent potential for error.

The problem, however, is more serious in pedagogical research. If we use a voltmeter, we may have to re-set the zero on the scale. If there is no scale on our meter, we may have to set a zero and calibrate the meter scale against known voltages. Our tests and examinations do not come with a zero defined at the outset and have never been calibrated. Indeed, we have a problem: how do we calibrate a test? The only way is to make a very large population of students sit the test and calibrate the marks so that marks in the test mean something to us. This is usually not feasible.

A 50% in an examination means nothing unless there is some external calibration of the test against some kind of known population. The key point to note is that marks must be treated with considerable caution. Our test may place the students in an approximate order of ability in some area of our course, but they have no *absolute* meaning whatsoever.

There is a third source of error which applies very strongly in pedagogical research. The marks an individual student gains in a test cannot be used as a basis for drawing conclusions. The student may be atypical or may have had a bad day. In all pedagogical work, we have to use large groups of students. To draw general conclusions, the larger the group the better because we even out the various anomalies of performance from individual students.

Thus, if we have 100 students who perform, on average 5% better in a test than another 100 students, we may be more confident in saying that the first group did better. However, the results could have emerged by chance. If, however, we compare two groups of 10 students and the difference in performance is 5%, then the possibilities that it happened by chance are much greater. This is where statistics can help us.

Use of Statistics

There are two main uses:

(a) To ***describe*** the information we have obtained:
For example, we can note the mean mark, the spread of the marks, the percentage of students who obtained the right answer to question 5, the proportion of students who are female and so on.

(b) To ***make deductions*** (or inferences) from the information we have obtained.
For example, we may wish to check if the average marks of two groups (which differ by, say, 5%) are likely to reflect a real difference in performance or are simply happening by chance (because of the errors of measurement or that the two groups were different in some other way)

Our Students

Let us suppose that we want to find out if some new approach in our teaching has improved student performance in examinations. We are unable to try out our new approach with half of the year group. Perhaps this is organisationally impossible, perhaps it will lead to potential student unrest, or we may feel it is unethical. However, we have the marks for the same course with last year's student group. Suppose we taught 85 students last year and this year's group contains 93 students. We can compare the marks if:

(a) We are reasonably confident that the two student groups do not differ too seriously in some way (for example, entry criteria were changed, the proportion of men/women has changed markedly, and so on).
(b) Last year's examination is the same as this year's. That is not so easy to achieve.
(c) The paper used in the examination is not designed in such a way that it favours one or other approach to teaching.
[For example, suppose we have introduced an element of problem-based learning into our course and wish to explore its benefits or otherwise on performance. If the examination only tests recall and application of information in a routine way, then any benefits which might arise from problem-based learning will probably not be seen.]

What statistics can do is to allow for the differences in the two group sizes, and to give us an indication in probability terms of the likelihood that any differences we observe in performance might have arisen by chance.

However, statistical manipulation cannot get round problems which are fundamental. We cannot compare the performance of two groups using two different measuring instruments (tests) unless the two tests are calibrated against each other (which may be impossible). We cannot compare the performance of two groups to see if our change in teaching has produced a better result if the two groups differ in some other way.

If our two groups are very small, then comparisons by this kind of approach are probably unhelpful anyway. A simple rule of thumb is to note that a sample of 200 gives the ideal use of most statistics but samples down to about 50 can still make statistical manipulation helpful. With smaller groups other approaches may be more helpful.

A Summary

(1) Look at the kind of information you obtain - that determines what you can do with it using any statistical approach.

(2) All measurements are open to error:
- (a) We can make errors in measurements;
- (b) Our instruments (tests) are uncalibrated;
- (c) Our student group may be too small and things can happen by chance.

(3) Statistics can be used to
- (a) Describe what we have observed (eg.. averages and totals);
- (b) Make deductions (inferences) (eg.. this group performed better than that group).

(4) The statistical method used must
- (a) Be consistent with the kind of data to be analysed;
- (b) Not make assumptions about the data which are not true.

More About Statistics

There are three fundamental questions which need to be explored at the outset:

(a) Do we need statistics at all ?
(b) What question(s) will statistics be able to answer ?
(c) What kind of data do we have ?

Let us take each in turn:

(a) **Do we need statistics at all ?**

Statistics can sometimes be used to give a kind of respectability to data which have been collected. However, any statistical manipulation can only be as good as the input data and there is a real danger in giving a false impression by invoking statistical techniques. No amount of sophisticated statistical manipulation can make sense or respectability of poor quality data or data which is prone to high error levels.

For much pedagogical research, very little statistical manipulation may be appropriate. Statistics should be seen as a tool to summarise and enlighten. Much research will need little statistical input.

Statistics can be used to summarise information (the average or mean of a set of marks, the standard deviation of these marks gives a measure of their spread). Such summaries may be useful but it is important to remember that such summaries are only as good as the data which we have collected. If we have 10 students who have sat a test, we can calculate a mean, a standard deviation and even the percentage of the class who are women. Such calculations are not very helpful simply because the class is so small. For example, one or two unusually able students will make the results misleading.

The real power of statistical manipulation lies in its ability to enable us to draw conclusions and make inferences. For example, in a test of ability to perceive three dimensional shapes in chemical molecules, 46 women and 54 men gained mean marks of 64% and 72% respectively. The questions is whether the different means indicate that men are better than women at this skill or could this result have arisen just by chance? A simple statistical calculation will give us information about the level of probability that such a result could have arisen by chance and is, therefore, not significant or is worthy of note.

There are some simple general principles: statistical calculations, conducted correctly, will give correct results. There is a view that statistics will prove anything. The abuse of statistics might be able to do this but the actual mathematics computed correctly leaves no room for doubt. However, the mathematics must not violate the nature of the data. ‘Soft’ data cannot be made into ‘hard’ data’ by statistics.

I promise to tell the truth, the whole truth, and nothing but the truth, 64.8% of the time.

(b) **What kind of question(s) will statistics be able to answer ?**

Much of this revolves round probability. The problem with pedagogical research is that we cannot draw conclusions from the performance of just a few students. This is exactly the same as medical research, for example. To check if a new drug improves recovery from some condition, no medical researcher would use only a few patients. The research would take a large number of patients. Perhaps the number would be divided into two groups, trying to make the groups as closely matched as possible. One group would be given the new drug while the other would be given an inert medication (a placebo), with neither group, nor their medical consultants, knowing who was getting which. The medical results with all the patients would be considered.

The question is whether the group receiving the new drug did better than the other group. Perhaps, an improvement in the cure rate could be attributed to chance. This is where statistics really comes into its own. The mathematics of probability is well established and it is relatively straightforward to take the data and check it statistically to see to what extent any improvement is attributable to chance.

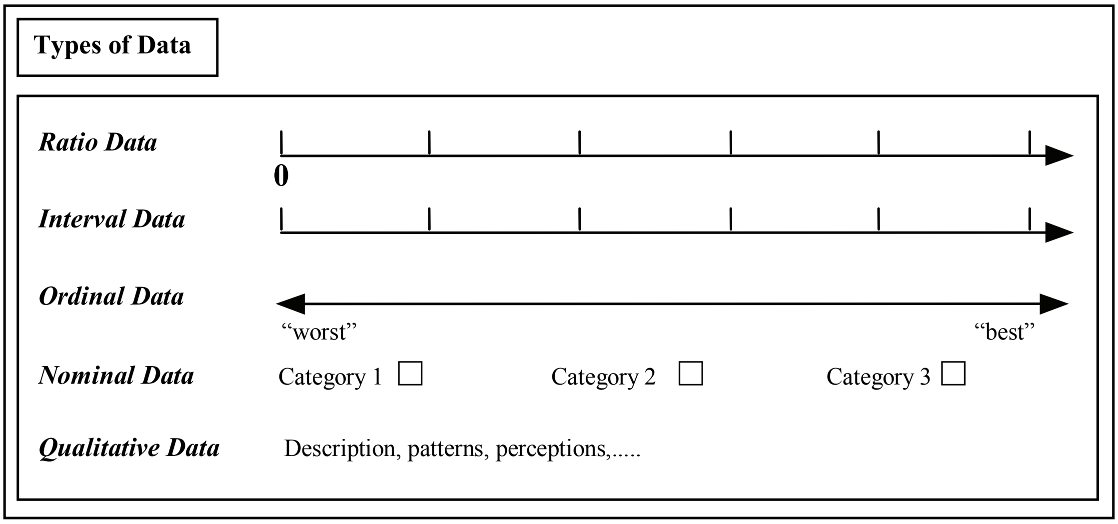
In pedagogical research, two broad questions will arise frequently.

(a) To what extent is some improvement in performance (in, say, a test) likely to have arisen by chance ?
(b) To what extent is some relationship in two different measurements likely to have arisen by chance ?

This second question needs some further explanation. Suppose you have noted the performance of your students in a mathematics test (say in differential and integral calculus) and you have their marks in a physics test (in, say, thermodynamics). It appears that students who do well on one test also do well in the other. Could this have arisen by chance? There is a statistical calculation (correlation) which is easy to do (given a computer) which will answer the question. It gives a numerical value for the correlation coefficient (from -1 to +1) and will tell you, for your group of students, the extent to which any pattern in performance is likely to have arisen by chance. It will ***not*** tell you if performance in mathematics is the *cause* of the physics performance or vice versa. It merely tells you if the two performances are linked, in the sense that those who do well in one test tend also to do well in the other. This might be nothing to do with the nature of the content of the tests but simply to do with the general ability of the students.

(c) **What kind of data do you have ?**

On page 43, we discussed several different types of data which can arise in pedagogical research. This was summarised in a diagram, shown here again for clarity:

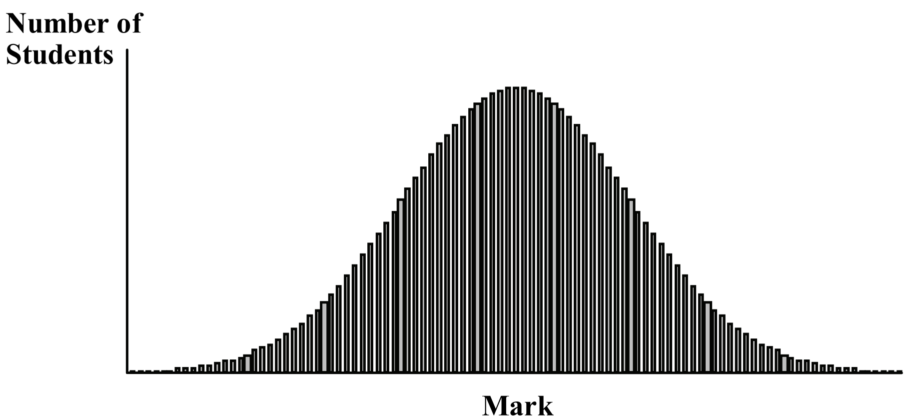


The mathematics behind various statistical techniques has been developed on the basis of certain assumptions. If we use a statistical technique which was developed on a set of assumptions which are not true for our data, then we may get misleading conclusions. One of the key divisions is:

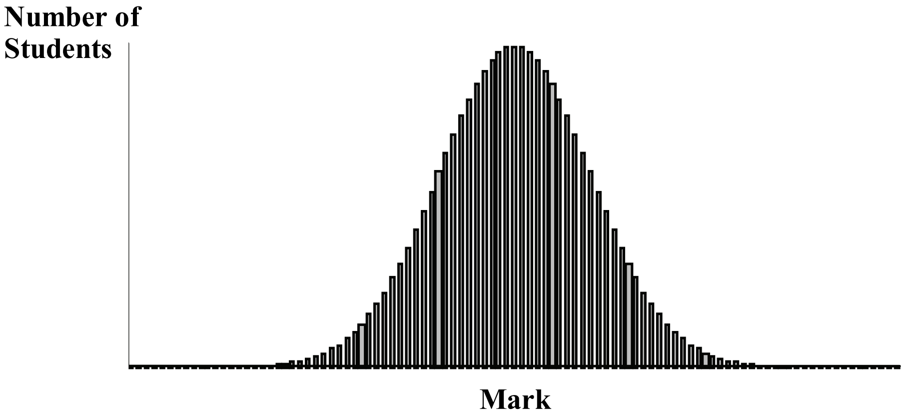
(a) ***Parametric data** - data drawn from a population of students where a defined distribution (usually a normal distribution) can be expected*

(b) ***Non-parametric data** - data drawn from a population of students where it is not possible to predict any particular distribution.*

If we set a test in a chemistry class, then we might expect the marks to show an approximately normal distribution: most students will bunch near the middle with some high-fliers and some poor performers. The larger the class, the closer the distribution might come to that of a normal distribution. The normal distribution looks like:



If the students were more bunched towards the mean mark, the distribution might look like:



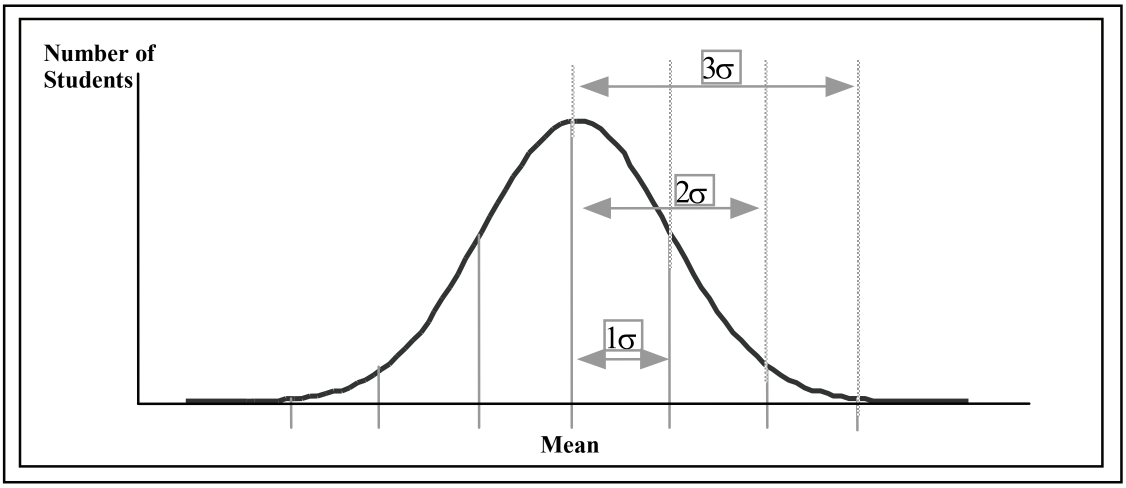
In a normal distribution, the ***mean*** specifies the average mark and the ***standard deviation*** specifies the mark spread around the mean. Means and standard deviations can vary, making the curve higher or flatter as well as moving up and down on the x-axis. However, the essential shape is always the same.

Most data based on tests and examinations tend to approximate to a normal distribution. Parametric statistics can be used as long as your data is not too wildly different from a normal distribution. The real problem occurs when your data is nothing like normal. This occurs, typically, in measurements of student attitudes when all kinds of distributions can occur. For example, if you ask 100 physics students what their attitudes were towards chemistry, it is possible that most would show negative attitudes! Equally, it is possible that attitudes might be polarised with some showing negative attitudes and some showing positive attitudes. With few students indicating a 'neutral' attitude, such a distribution of attitudes would be completely different from a normal distribution. The use of parametric statistics would be completely inappropriate.

As a general rule, because of the assumptions which are made, parametric statistical tests are more sensitive but they may give misleading results if used on data which is far from normal in distribution. Then, the non-parametric tests must be used.

More About Probability

Some features of the normal distribution curve are shown in the following diagram below.



The standard deviation is defined as the *distance from the mean to the first turning point of the curve* (shown as sigma, σ). It turns out that 68% of the area under the curve lies between 1 standard deviation above the mean and 1 standard deviation below. That means that, if we have our chemistry marks forming a perfectly normal distribution, 68% of the students will gain marks between 1 standard deviation above and 1 standard deviation below the mean.

It is also true that 95% of the student population will have marks which lie between 2 standard deviations above and 2 standard deviations below the mean.

> Let us suppose that, in our chemistry test, we have a mean of 60 and a standard deviation of 10. Then, we expect to find that 95% of our students will have mark between 40% and 80%. For any given student, there is only a 5% chance that they will gain a mark higher than 80% or lower than 40%. It is also worth noting that 99% of our students will have marks which lie between 30% and 90% (the positions of three standard deviations below and above the mean).

If our statistical calculations are based on mathematics which assume the mathematics of a normal distribution, we have the beginnings of being able to see probabilities of measurements and comparisons being likely to have arisen by chance or unlikely to have arisen by chance. We can also specify the probabilities that a result is likely to have arisen by chance.

In many computing packages which carry out the calculations for us, probability is specified in terms of 5% probability or 1% probability. For example, a result is said to have a probability of less than 1%. We can say that we are more than 99% sure that our result did ***not*** arise simply by chance.

Just to confuse things, there are two ways by which probability is expressed in statistical packages:

(a) ***Using percentages****:* *e.g. the results are significant at <5%. This means that the chances of our result occurring simply by chance is less than 5%.*

(b) ***Using ratios****:* *e.g. the results are significant at < 0.05. This has exactly the same meaning, the scale of probability being related to 1 rather than 100.*

Both systems are in use.

Interpreting and Applying Our Results

The purpose of pedagogical research is to offer understandings of the processes of teaching and learning so that student learning may become more effective and more efficient. How can we be sure that the conclusions we draw from our observations are, in fact, true for the learners who took part? Even more important, are these conclusions likely to be true for others?

This section explores such questions.

Are we measuring what we think we are measuring ?

In exploring teaching and learning, data have to be collected. This can be done by tests, questionnaires, interviews, observation. Are the data we collect giving us information of the kind we expect. This is often described as validity: are we measuring what we think we are measuring? To check this, we need to seek some kind of criterion external to the test itself. This may involve some completely separate evidence or it may rely on the views of others who are competent to judge our testing procedures.

Example:

In a survey by questionnaire of the factors which influenced university students towards a degree in physics, students indicated that the nature of the topics to be studied was important. They seemed to indicate that they were attracted towards themes and topics which they saw applying to their life style. To check if this was a valid observation, a school syllabus was found which had been constructed in such an applications-led way. It was observed that the proportion of students who chose to continue with physics studies after completing this syllabus was higher than the retention rate for any other school subject at that level. Here was external evidence that the observations from the questionnaire were, in fact, likely to be valid.

This leads to a general principle in all research. We shall always have more confidence in our conclusions if they are confirmed by some completely separate source of evidence. This kind of approach is sometimes described as *triangulation*. However, validity of any form in research is elusive; we can never be sure of having achieved it.

Sometimes we have to rely on the expert opinions of colleagues. We do this regularly when we set tests and examinations. Colleagues will offer constructive advice and criticism about the test questions we set, this being offered from a perspective of experience. In conducting pedagogical research, the constructive advice and criticism of colleagues is often vital and may constitute clear evidence for the validity of our measurements.

Can we rely on our measurements?

In making measurements in the physical sciences, we may assume that a measurement we make one day is likely to give the same results when we repeat it the following day under similar conditions. We also assume that the measurement itself is not altering the thing which is being measured. Neither assumption will always be perfectly true but we can have a fair idea if the errors are likely to be serious.

The same principles apply in pedagogical research. We want to know if our measurements are reliable and so will give the same results on separate occasions. Educational researchers have come up with all kinds of ingenious statistical approaches in seeking an answer to this question and most of them are ***not*** very helpful. Some examples illustrate the kinds of approaches which may prove useful.

> ***Example***
>
> *In an experiment to see if some practical work offered helpful insights into solving a problem in chemistry, the researchers tested a sample of students with the problem. They then allowed the students to watch and interact with an experimental situation. A day or two later, they then re-tested the students on the same problem, looking for an improvement. They were fully aware that the first time the students tried the problem (which was difficult for them) may itself have brought about an improvement in attempting the problem a few days later. In fact, there was little improvement in the students' problem solving success and they were able to conclude that the experimental input was not very helpful. They carried out further experimentation to explore this further.*

In essence, if a length is to be measured, then, if two measurements give the same or very similar results, we are satisfied. It assumes that the object to be measured is not affected by the actual measurement and the object has not changed in size. In practice, time rarely permits us to make repeated measurements like this in pedagogical research. However, if our tests are of reasonable length, of appropriate difficulty, avoid verbal ambiguity and are applied under appropriate test conditions, it is likely that our measurements will be reliable.

Most statistical measures of reliability only give evidence about test consistency. However, many tests and questionnaires have questions which are deliberately ***not*** designed to measure the same thing. Consistency across questions is, therefore, meaningless. The problem lies in the way educational research has drawn from psychological research where internal consistency is often extremely important. Most statistical measures of reliability have little applicability in pedagogical research.

Advice: An invalid test may still be reliably useless! Beware reliability tests!

Our Overall Aim

In an ideal world, we must be seeking for two pieces of evidence, gained independently of each other, which point the same way and thus validate each other. We must explore potential rival interpretations and look for the exceptions or negative cases. These exceptions often are critical in seeing the way specific conditions may apply to our conclusions.

In carrying out pedagogical research, you may be involved on your own as teacher. Your sample may inevitably be limited (in size, institution, context etc.) and the opportunities to carry out experiments to check your conclusions may be limited. Ideally, you are looking for an approach to pedagogical research which allows you to formulate hypotheses and develop strategies applicable to the teaching situations in general. There are two ways to approach this:

(a) Look at the literature - see, for example: <www.heacademy.ac.uk/physsci/publications/practiceguides>
(b) Engage colleagues in parallel work - within and beyond your institution
(c) See your work as ongoing. Repeat next year, with improvements or extensions

Other Kinds of Research Evidence

There are several other types of research. Being sure that the data we obtain is measuring what we think it is measuring and is repeatable is not always easy:

(a) ***Open-ended Surveys***: Sometimes, a research programme has to start with an initial survey simply to fund out what is happening.

> Thus, for example, suppose you have just taught a new course. You wish to find out where students found difficulties. There are two possible sources of evidence you are considering. You can list all the topics and themes of the course and ask the students to indicate which caused them problems but how can you be sure that you can regard their answers as honest or correct? After all, they may not be aware that they do not really understand a particular topic. Of course, you could always look in detail at their examination performance but how can you be sure that poor performance reflects a real difficulty rather than an over-demanding question, perhaps badly set?

If you compare the data from your survey and the examination, then where both indicate difficulty, you can be fairly confident that you are observing a real problem area. Where there are inconsistencies, this defines the agenda for the next piece of research! As an aside, such surveys of difficulties usually are fairly accurate in pinpointing most of the problem areas.

(b) ***Testing an Hypothesis***. Sometimes, you will be in a position to test an hypothesis which has arisen from previous research work.

> In a piece of work which was looking at physics laboratories, an Information Processing Model predicted that student performance (in terms of understanding the physics) would increase if the information load could be reduced. The researchers had looked at the sources of information overload and appreciated that pre-laboratory exercises would be expected to achieve this load reduction. In an ingenious experiment, they tested the achievement by two different methods (demonstrator marking and post-laboratory exercises which tested understanding) and found that data from both sources confirmed quite considerable increases in student learning.

With two independent sources of data, the researcher could be very confident that they were observing a real increase in learning. For more details of this particular piece of research, see page 58.

Aiming at Quality

This section looks at the sharing of pedagogical research and how research must be conducted in a way acceptable to our students and others.

The aim is quality and some characteristics of quality are described.

Encouragement to carry out pedagogical research is given.

Making Your Findings Known

One of the sad features of life in higher education is that there is little credit given for commitment to teaching when compared to research in the physical sciences. One way to seek to emphasise the importance of teaching is to focus some effort on publishing outcomes from pedagogical research. Fortunately, there has been a growth in journals and there are greater opportunities for refereed publication in quality publications.

Some journals focus strongly on the exchange of useful information and experience with regard to teaching and learning but they do take some papers based on pedagogical research. Examples include Physics Education (IOP) and Education in Chemistry (RSC) in the UK and the Journal of Chemical Education and the American Journal of Physics in the USA. These journals do not demand major literature reviews from the educational literature but seem to be more interested in 'things' that seem to work and are useful, with some evidence.

The RSC also has University Chemistry Education which has a high regard for pedagogical research standards as well as being directed specifically towards Higher Education. There are numerous education journals and it is worth checking their editorial policies. Some may be useful outlets. The International Journal of Science Education, Research in Science and Technological Education, the Journal of Science Education (Spanish/English) and Chemistry Education Research and Practice (CERP) are all international journals of quality which may be considered. In these journals, there is an expectation (in varying degrees) that there is some coverage of the relevant educational literature.

Although there are variations from journal to journal, the format of papers will contain a statement of what you were trying to do and why, a summary of the relevant background literature on the theme to demonstrate that your work is set in the context of previous work, an outline of your methodology, a discussion of the results obtained, with conclusions and suggestions about applications to teaching and learning and future research.

A useful way to make your work known to the group of colleagues throughout the UK who are likely to be most interested in what you have done is to offer material for the Physical Sciences Centre publications. In addition, university libraries are now taking practical steps to publish electronically what they call 'grey' literature. This may prove a valuable way to exchange information without the time-consuming formalities of developing publications for refereed journals.

Guidelines on Ethical Research Practice

Good research is not just about the quality or even the quantity of data gathered but also the way in which the work is undertaken. Most higher education institutions will have their own policies on ethical issues, but here is a very brief summary of the main issues when working with your own students.

(1) Responsibilities and Relationships with Participants

(1) Research should be based on informed consent.
(2) The students should be given anonymity and confidentiality unless they express otherwise.
(3) It is important to take care that any information gathered is accurate. This may involve showing and/or sharing this information with the participants.
(4) Participants have the right to feedback on research findings.
(5) If outcomes from the research may be placed in the public domain (e.g. publications, conferences,) it is essential to maintain complete anonymity.

(2) Practical Aspects in Working With School Pupils

(1) The usual procedure is to seek permission from the local authority to approach certain named schools and then, if permission is granted, to contact the head teacher of each school seeking permission to gain access to pupils and/or staff. Some Local Authorities also require parental consent. It is important to set out in writing what access is being requested, for how long and with what purpose, following the guidelines above. It is important to recognise that such procedures can take months.
(2) In visiting schools, it is essential to make the approach through the school office, following the procedures of the school with regard to security.
(3) As a matter of courtesy, a brief letter of appreciation (or phone call) should be sent to the key personnel (head teacher, head of department or teacher) involved after access to a school and it is always useful to send a summary of any outcomes of the work completed.

(3) Ensuring Anonymity and Confidentiality

(1) Anonymity means that information cannot be traced back to individuals. Special care must be taken with interview data.
(2) Risk minimisation must be applied to all research data. It should be kept confidential and should be stored in a secure manner. Where possible, use codes or other identifiers to break obvious connections between data and individuals. Where there is a mixture of information which is in the public domain and that which has been obtained by informed consent which concerns the same organisation, no traceable link must be left.
(3) When placing the data, or the findings derived from the research into the public domain, it is important to remove any identifier which could be traced back to the participants in the study. The 1998 Data Protection Act has specific requirements.

(4) Useful Links to other Guidelines on the Ethics of Educational Research

(a) BERA: "Ethical Guidelines" http://www.bera.ac.uk/guidelines.html
(b) The Data Protection Act 1998 http://www.hmso.gov.uk/acts/acts1998/19980029.htm.
(c) JISC (Joint Information Systems Committee): "Guidelines on Policy Approaches to IPR in HEIs"
http://www.strath.ac.uk/ces/projects/jiscipr/

Characteristics of Quality Pedagogical Research

It is difficult to specify unambiguously the criteria for quality pedagogical research in that there are so many different areas with different emphases.

Quality Research

- **Clear statement of aims and objectives (a question worth exploring)**
- **Awareness of research context (literature, potential significance)**
- **A robust methodology (eg good samples, good test materials, control of variables)**
- **Appropriate data handling (eg careful and appropriate use of statistics)**
- **Results which can be interpreted unambiguously (based on good experimental structures)**
- **Conclusions drawn appropriately and applied appropriately**

Excellent scientific research often involves well-led teams of researchers who, over time, carry out many projects where the outcomes from earlier projects point to the agendas for the later projects. Pedagogical research is essentially the same. The most useful work rarely comes from one-off experiments.

In developing pedagogical research, therefore, let your first (perhaps small) project define some of the agenda for the next project and so on. Seek to work with colleagues or research students to develop a coherent programme which, over time, gives a pattern of results which can be really helpful and useful.

Quality research cannot be learned from a textbook nor even by reading this manual! We learned how to plan and carry out meaningful experiments in the physical sciences by planning and carrying out experiments in the physical sciences. We all learned from our mistakes and each experimental experience gave us the confidence to move on to try something new and more ambitious. It is exactly the same with pedagogical research.

General Principle

The best way to learn how to conduct pedagogical research is to conduct pedagogical research.

Summary of Some Research Projects
(from the literature)

Illustrating Various Approaches to Pedagogical Research

This section outlines twelve pedagogical research studies from the literature.

They have been chosen to illustrate some the approaches which have been used successfully in the recent past.

They offer models and exemplars of quality pedagogical research in the physical sciences.

(1) Conceptual Understandings

Undergraduate Students' Understandings of Enthalpy Change
University Chemistry Education, (1999)*,* **3***(2), 46-51*

The authors describe a study where they were seeking to address the issue relating to student understanding of basic concepts in thermodynamics. This can be distinguished from the ability to use mathematical relationships correctly to gain correct answers. They interviewed 20 students before and after a first year course in thermodynamics, using three demonstrated experiments as the basis of interview. They list the key concepts being explored and note the success rate in understanding in the two sets of interviews and also discuss briefly some of the comments and answers given by the students.

This paper illustrates the way interviewing can be used imaginatively to generate useful insights into student learning. It also illustrates one way by which data from interviews can be summarised and presented so that the impact of the research evidence is accessible to the reader. Another feature of the paper is the way their work is related usefully to other studies. In passing, their evidence is frightening as it confirms much of what many of us have feared for some time.

(2) Lectures

Lectures - a Learning Experience ?
Education in Chemistry (1994), **31**(3), 75-79.

This is a very compressed account of a series of major experiments which looked at the lecture experience. The researchers had already measured the working memory capacity of a group of 75 students, a straightforward and quick process. They had also measured the students' field dependency (this is a measure of the students' ability to distinguish the 'message' from the 'noise' and, again, is measured in a straightforward way). They then analysed a long series of lectures, given by many lecturers, in terms of the numbers of words spoken, the number of ideas presented, the amount written on a blackboard (or equivalent), the recorded student notes both in terms of quantity and nature, the student performance in examinations.

This was clearly a long and tedious piece of work, carefully conducted, which showed very clearly the characteristics of lecturing as a mode of instruction, the characteristics of the effective lecturer and the way that the minority of students who happen to have innate advantages in the capacity of their working memory and ability to distinguish the 'message' from the 'noise' have huge advantages in learning by means of lectures. It illustrates how many sources of data can be combined to give an overall view which is full of useful insight as well as raising matters of considerable concern.

(3) Physics Laboratories

The Students' Attitude and Cognition Change to a Physics Laboratory
Physics Education, (1998), **33**(1), 22-29.

This paper offers an excellent highly accessible introduction to a more pedagogical way of thinking, linking the realities of physics teaching to well established educational ideas. This paper illustrates the careful way in which the researchers controlled as many variables as they could, both by scrupulous care in working with staff and demonstrators but also by the experimental set-up.

In their experiment with about 180 students, the researchers had modified the physics laboratory experience so that experiments had pre-laboratory exercises to be completed in advance. Each student completed two experiments with pre-labs exercises and post-labs exercises and two experiments with no pre-labs exercises but with post-labs exercises. Different students completed different combinations of experiments, thus removing possible variation between student samples. The pre-labs exercises and post-labs exercises were designed with a clear educational model in mind.

The attitudes of students to each experiment were measured using a Likert-based approach. The remarkable observation is that quite incredible attitude differences were observed. The researchers also attempted to measure the gain in learning arising from each experiment in the laboratory by means of a standard demonstrator marking system and also by marking the post-lab exercises. Both measurements showed quite remarkable gains in learning arising from the use of the pre-labs exercises.

One of the interesting and unusual features of this piece of research is the way that learning and attitudes are linked. This paper demonstrates a very good experimental set-up for variable control and the gaining of meaningful data.

(4) Chemistry Laboratories

An Information Processing Model of Learning: its application to an undergraduate laboratory course in chemistry.
Studies in Higher Education (1994) **19**(1), 77-87.

This paper illustrates research which used a set of four hypotheses which had been predicted by a well-established model of learning. An inorganic laboratory was re-designed to test these hypotheses. It is an example of a systematic, theory-led approach, illustrating the value of educational models of learning when applied in a systematic way.

The laboratory was re-designed, introducing pre-labs exercises and mini-projects. The paper illustrates a nice research model where, out of five groups of students, two were control groups and three were experimental groups:

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
	Control	Pre-lab	Mini-project only	Pre-lab	Control
Improved manual	Yes	Yes	Yes	Yes	Yes
Pre-lab	No	Yes	No	Yes	No
Mini-project	No	No	Yes	Yes	No

This piece of research also illustrates some ingenious ways to obtain information about the success or otherwise of the various laboratory approaches. Their aim was to relate student attitudes and outlooks to the various laboratory arrangements. Apart from the usual questionnaires, the researchers looked at student diaries, demonstrator diaries and demonstrator checklists. Together, all these sources of data give a rich picture of what was going on, showing very clearly that pre-labs exercises plus mini-projects was very much the best way to run such a laboratory.

(5) The Pre-Lecture Idea

Preparing the Mind of the Learner
University Chemistry Education (2001), **5**, 52-58.

This study illustrates what can be revealed by studying examination performance over a period of time when specific key changes are introduced into the way a course is offered.

In a large class, pre-lectures of a specific kind were introduced for two successive year groups. The effect of these on the *least well qualified in chemistry* was marked. For the following three years, the pre-lectures were not offered and the least well qualified showed a marked drop in performance. In the sixth year, an attempt was made to develop teaching materials which tried to mimic the aims of pre-lectures. A similar positive effect on the performance of the least well qualified in chemistry was again observed.

The particularly interesting feature of this work lies in the way a well established educational model of learning was employed to underpin the design of curriculum materials and the way that this design brought about a performance enhancement which benefited the least well qualified students most, a most unusual observation. Another unusual feature is that this paper describes a study which spanned 6 years of students groups.

(6) Attitudes to Physics

Attitudes Towards Physics
Research in Science and Technological Education, (2002), **20**(1), 67-81.
Gender and Physics
International Journal of Science Education, (2003), **25**(4), 509-536
These two papers give a detailed account of a major study which looked at attitudes of learners to physics from age 10 to age 20 (from the sixth year of Primary to the second level physics students at a university). Inevitably, most of the papers concentrate on school pupils but they illustrate the methodology where questionnaires and interviews were conducted with several different groups of learners of different ages and stages. Comparisons were then made between the data obtained for the age groups. This offered on overview of attitude development which was able to pinpoint the key issues affecting learner attitudes to physics.

The research did not rely on more traditional approaches to attitude measurement (with all their inherent logical and statistic flaws) but used a wide range of techniques. It is also worth noting that this work arose because of concerns about retention rates at university in physics but revealed that, at least in Scottish schools, the retention rates were higher for physics than any other elective subject. Because of its size and scope, this study illustrates how clear patterns and findings can be observed which offer direct answers to the problems often associated with the popularity of physics.

(7) Physics Problem Solving

A Working Memory Model Applied to Physics Problem Solving
International Journal of Science Education, (1993), **15**(6), 663-672.

This is an early study on problem solving where an educational model was used to make sense of the data obtained. Although the paper is brief and many ideas are compressed, it is written extremely clearly. In essence, problem solving success was related very remarkably to the innate ability of the students to handle information. Each of us has a working memory space in our brains which is of limited capacity. It is used to hold and manipulate information when solving a problem. If a physics problem requires more working memory space for its solution than a student possesses, failure is more or less guaranteed. The paper also analyses the students' ability in separating the 'message' from the noise'. The study is highly quantitative, with remarkably precise conclusions.

There were two main experiments described:

(1) The researchers set a test with questions of variable working memory demand, this demand being defined as the least number of steps (amount of information) to be handled to reach a solution. The results, as expected showed that as demand increased, success fell. However, the fall was not linear but dropped cataclysmically as soon as the demand exceeded the working memory of each student (which was measured separately). This is a useful example of a study where a developing working model of learning was used to design an experiment, the results of which turned out to be very clear-cut.

(2) The researchers developed a set of physics questions, each of which was offered in several formats:

(a) Question set as text only;
(b) Same question set in diagram form;
(c) Same question set using both text and diagram;
(d) Same question in a simplified form with more information given;
(e) Same question in a highly structured form.

This is a nice example of the control of variables so that conclusions can be drawn. The experiment was developed in the context of a predictive educational model.

(8) Misconceptions

A Report of Undergraduates' Bonding Misconceptions
International Journal of Science Education, (2001), **23**(7), 707-730.

This paper describes a study of misconceptions at various stages related to certain topics in chemistry. It illustrates the use of hour long semi-structured interviews and offers (as an appendix) an outline of the interview structure used. Table 1 in the paper reveals the enormous range of misconceptions and confused ideas while table 2 shows how many of the sample of 56 revealed each of the areas of difficulty.

Interviews normally generate qualitative data and, indeed, such data can offer very rich insights into student thinking. In this study, there were enough students to be able to offer a way to analyse interview data to gain some kind of quantitative data. Where you have the time to interview large numbers of students (using, perhaps, a team of interviewers), here is a suggested approach which might offer useful outcomes.

Again, although the topic is in chemistry, the approach is equally valid for any subject area.

(9) Student Understandings

Students can pass tests and examinations without necessarily having a good understanding of the material involved. To achieve this, they may depend on memorisation or on mastery in carrying out set procedures (like calculations)

In a long series of research projects, one research centre in the USA has explored this problem in relation to astronomy and physics. Three papers are summarised here and illustrate a sustained research programme and the way the researchers use test materials and interviews in diverse ways to illuminate the problems. The minimum of statistical analysis is present.

Student Understanding of the Wave Nature of Matter: Diffraction and Interference of Particles.
American Journal of Physics, (2000), **68**(Supplement 2), S42-S51

Students in first, second and third year physics courses were asked to predict and explain how a single change in an experimental set-up would affect the pattern produced when electrons or other particles were incident on a single slit, a double slit or a crystal lattice. They used written tests and interviews to gain insights into student misunderstandings. They describe the main problems they detected at each level and discuss their implications for teaching.

Student Understanding of the First Law of Thermodynamics Relating Work to the Adiabatic Expansion of an Ideal Gas.
American Journal of Physics, (2002), **70**(2), 137-148.

This study looked at students on a first year course and on a second year thermal physics course. The enquiry started by interviews where each student observed a demonstration and then was interviewed about its interpretation. Written test materials followed. This illustrates a use of interviews to identify areas of potential difficulty (defining the agenda) and shows very clearly how student understanding was very weak despite the lecture courses they had attended.

The Challenge of Changing Deeply held Student Beliefs about the Relativity of Simultaneity.
American Journal of Physics, (2002), **70**(12), 1238-1248

The researchers have noted that students construct incorrect conceptual frameworks and that these wrong ideas are very difficult to dislodge later. They developed teaching materials which attempted to correct such misunderstandings and misconceptions. In this study, they tested the students before and after the use of the new materials to gain information about improvements. They used various forms of test items so that no student faced the same items before and after. Their results are presented simply in terms of percentages of those who were successful, showing the benefits of the materials.

Statistics Appendix 1

How to Handle Marks

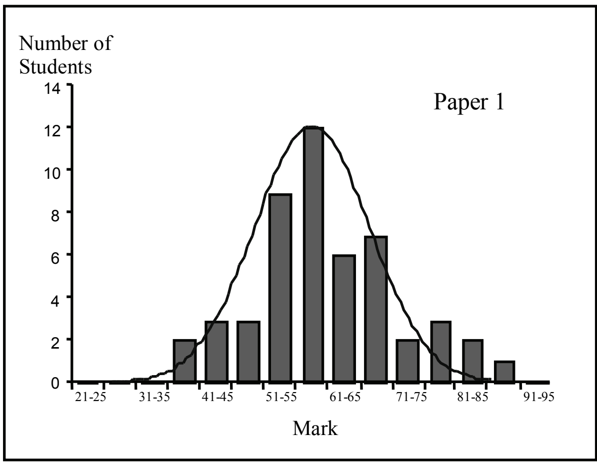
To illustrate the way data can be handled, imagine you have a class of 50 students (25 male, 25 female). Suppose they sit two examination papers (each marked out of 100) and a project (also marked out of 100).

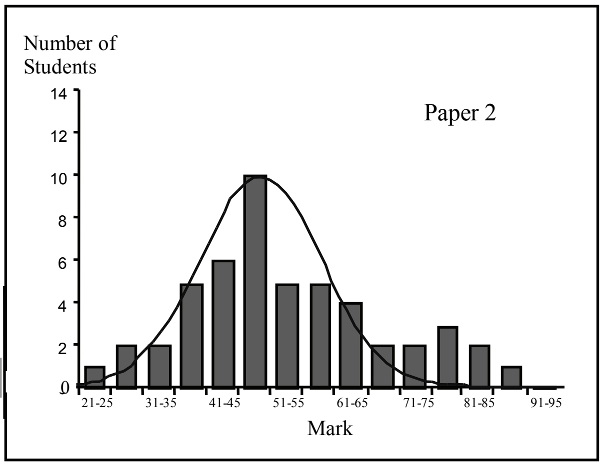
The marks are shown on the spreadsheet opposite.

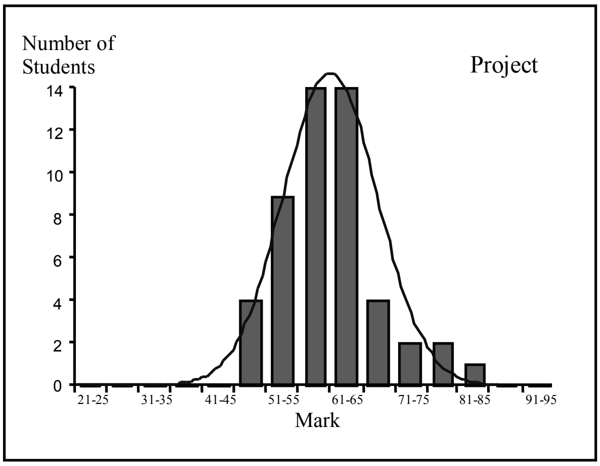
Here are some questions which could usefully be asked.

(1) **Do the sets of marks form any thing like normal distributions ?**

Most spreadsheets will calculate means and standard deviations and will offer graph drawing facilities. For example, Microsoft Excel or Appleworks (both work on both Macs and PCs) can be used. The following graphs were obtained from the marks data:







For Paper 1:	Mean Mark = 60 Standard Deviation = 11.1
For Paper 2:	Mean Mark = 53 Standard Deviation = 15.1
For the Project:	Mean Mark = 61 Standard Deviation = 7.8

All three distributions are reasonable approximations to a normal distribution and parametric statistics can be applied without too much error.

Paper 2 is more demanding than paper 1 and there is a long 'tail' at the upper end as the more able perform well.

The project is marked in a more narrow range, with few markers willing to give low marks.

The mark distributions are typical of the kind of marks you may obtain in your own examinations and show that, to a reasonable approximation, normal distributions are obtained, allowing the use of parametric tests without too much error.

The Student Marks

	Student Marks (all as %)			
Student	**Gender**	**Paper 1**	**Paper 2**	**Project**
1	M	68	58	63
2	M	54	49	61
3	F	41	30	52
4	M	55	49	80
5	F	68	60	52
6	F	57	40	55
7	M	59	50	60
8	M	81	88	59
9	F	44	36	58
10	M	35	27	50
11	F	55	50	65
12	F	68	61	60
13	M	72	59	60
14	F	44	32	69
15	M	54	43	54
16	F	82	81	64
17	F	37	25	48
18	F	61	45	61
19	M	66	58	63
20	M	60	57	62
21	F	56	44	54
22	M	62	51	63
23	F	56	52	59
24	F	58	47	57
25	M	61	68	60
26	M	66	78	50
27	M	48	33	66
28	F	65	51	63
29	M	78	62	56
30	M	73	77	64
31	F	54	37	56
32	F	60	77	66
33	M	87	75	78
34	F	56	36	54
35	M	48	41	59
36	F	61	63	72
37	M	77	72	70
38	M	59	39	52
39	F	70	67	50
40	F	55	48	59
41	F	55	42	52
42	M	59	50	63
43	F	67	50	55
44	F	58	46	73
45	M	53	48	57
46	F	48	44	63
47	M	59	52	64
48	M	62	81	58
49	F	57	63	65
50	M	76	51	85
Means		**60**	**53**	**61**
Standard Deviations		**11.1**	**15.1**	**7.8**

(2) **Is it fair to add the marks together to gain a score for the whole course ?**

Let us assume that Paper 1, Paper 2 and the Project have equal weighting. If we simply add the three marks together and divide by three, we obtain a rank order which may ***not*** reflect the measured ability of the students. This happens because the three sets of marks have different means and standard deviations. For example, a high mark in the easiest paper can outweigh a high mark in the most difficult paper.

To gain the most fair order of merit, the marks in the three assessments must be adjusted so that each assessment has the *same mean* and the *same standard deviation*. This is a little like adding up voltage drops around a circuit. We cannot add millivolts to volts directly. We must convert all voltage drops to the same unit, either all to volts or all to millivolts. In a similar way, we are taking our assessment marks and getting them all into the same units so that a mark in one paper means the same as a mark in the others. This is known as ***standardising marks***.

Fortunately, the arithmetic of this is easy and any spreadsheet can do this very easily (see below). The spreadsheet opposite shows what happens when the marks on each test are standardised. We have chosen a mean of 60 and a standard deviation of 12. The order of merit in any test is ***NOT*** altered by standardising the marks. However, we can now add the marks on the three tests together because they are now measuring performance on the same scale.

The final column shows the students total performance in the three tests. The standardised marks of the three tests have been added up and the total divided by three. This will give the fairest order of merit because the three measurements (tests) have been calibrated against each other onto the same scale. This fundamental principle is important in comparing any data. We assume that our measuring devices in the physical world all work to the same scale. A millivolt division on each meter is regarded as the same to within a very small error limit and, therefore, values from different meters can be compared, added, subtracted and so on. Educational tests are exactly the same except that we *choose* the average (mean) and the scale (standard deviation).

Of course, you can make the mean high or low and give the standard deviation a wide range of values. However, a useful mean is 60, with a standard deviation of 12. This means that approximately 68% of your students will gain a total mark between 48 and 72, and 95% of them will have a total mark between 36 and 84.

Standardising Marks

To convert marks in an assessment into standardised marks:

(a) Find the mean and standard deviation of the marks in your assessment (using a spreadsheet)
In Excel or Appleworks: Mean = AVERAGE(array on spreadsheet) and Standard deviation = STDEV(array on spreadsheet)

(b) Standardised Mark =
{[(Raw mark - Mean of your marks) ÷ standard deviation] x chosen standard deviation} + chosen mean.

(c) For example: a student has a mark of 65 in an assessment where the mean is 50 and the standard deviation is 15. Suppose you wish to convert this to a standard mark where the mean is 60 and the standard deviation is 10.

Standardised mark = {[(65 - 50)/15] x 10 } + 60 = 70

NOTICE: Standardising marks does ***not alter the order of merit in a paper***. It brings the marks of all papers on to the same scale where ***one mark in one paper is the same as one mark in another paper***

Standardising Test Scores

Student	Gender	Paper 1	Paper 2	Project	Paper 1	Paper 2	Project	Average Mark
					as Standard Scores	as Standard Scores	as Standard Scores	using standard scores for the tests
1	M	68	58	63	69	64	63	65
2	M	54	49	61	53	57	60	57
3	F	41	30	52	39	42	46	43
4	M	55	49	80	55	57	90	67
5	F	68	60	52	69	66	46	60
6	F	57	40	55	57	50	51	53
7	M	59	50	60	59	58	59	58
8	M	81	88	59	83	88	57	76
9	F	44	36	58	43	47	56	48
10	M	35	27	50	33	39	43	39
11	F	55	50	65	55	58	67	60
12	F	68	61	60	69	66	59	65
13	M	72	59	60	73	65	59	65
14	F	44	32	69	43	43	73	53
15	M	54	43	54	53	52	50	52
16	F	82	81	64	84	82	65	77
17	F	37	25	48	35	38	40	38
18	F	61	45	61	61	54	60	58
19	M	66	58	63	66	64	63	65
20	M	60	57	62	60	63	62	62
21	F	56	44	54	56	53	50	53
22	M	62	51	63	62	59	63	61
23	F	56	52	59	56	59	57	57
24	F	58	47	57	58	55	54	56
25	M	61	68	60	61	72	59	64
26	M	66	78	50	66	80	43	63
27	M	48	33	66	47	44	68	53
28	F	65	51	63	65	59	63	62
29	M	78	62	56	79	67	53	66
30	M	73	77	64	74	79	65	73
31	F	54	37	56	53	47	53	51
32	F	60	77	66	60	79	68	69
33	M	87	75	78	89	78	87	84
34	F	56	36	54	56	47	50	51
35	M	48	41	59	47	51	57	52
36	F	61	63	72	61	68	77	69
37	M	77	72	70	78	75	74	76
38	M	59	39	52	59	49	46	51
39	F	70	67	50	71	71	43	62
40	F	55	48	59	55	56	57	56
41	F	55	42	52	55	51	46	51
42	M	59	50	63	59	58	63	60
43	F	67	50	55	67	58	51	59
44	F	58	46	73	58	55	79	64
45	M	53	48	57	52	56	54	54
46	F	48	44	63	47	53	63	54
47	M	59	52	64	59	59	65	61
48	M	62	81	58	62	82	56	67
49	F	57	63	65	57	68	67	64
50	M	76	51	85	77	59	97	78
Means		**60.1**	**52.9**	**60.8**	**60.0**	**60.0**	**60.0**	**60.0**
Standard Deviations		**11.1**	**15.1**	**7.8**	**12.0**	**12.0**	**12.0**	**9.6**

(3) **Do the males do as well as the females in each assessment?**

The spreadsheet will quickly calculate the average mark for the 25 males and 25 females separately and the standard deviations of their marks.

	Paper 1	Paper 2	Project
female mean	57.3	49.1	59.3
male mean	62.9	56.6	62.3
female standard deviation	9.9	14.0	6.7
male standard deviation	11.8	15.4	8.6

This set of data is not easy to interpret and this is where we need statistical help. Because the distributions are not too far off normal, the correct statistical test is the ***t-test***. This can take account of different means and standard deviations and will tell us the probability that the males and female performances arose by chance. Although a spreadsheet like Excel can do the calculations, it is much better to use a statistical package like SPSS (another package is Minitab). The data can be pasted into SPSS (which looks rather like a spreadsheet).

The best way is to use a handbook which guides you through the use of SPSS[9]. In SPSS, there are several choices offered when comparing Means (found under the Statistics pull-down menu). In this case, select independent samples.

The results are:

Test 1	$t = 1.81$	$p = 0.078$	not significant
Test 2	$t = 1.81$	$p = 0.076$	not significant
Project	$t = 1.38$	$p = 0.175$	not significant

The key figures are the probability figures which SPSS quotes as fractions of 1.

Thus, a probability given by $p = 0.078$ means that the probability that the differences were caused by chance is 7.8%. Usually, if the probability is more than 5% ($p > 0.05$) then the result is considered as *not significant*: we consider that it is likely to have arisen by chance. Typically, pedagogical research looks for probabilities of 5%, 1% and 0.1% ($p<0.05$, $p<0.01$, $p<0.001$).

Thus, in all three tests, we conclude that any observed differences between men and women are not significant and were caused by chance. Looking at the data in the table above might have led us to conclude otherwise. However, the small sample (25 men and 25 women) make it difficult to conclude that there are gender differences in performance.

9 Gray, Colin D., and Kinnear, Paul R., (1998)) SPSS for Macintosh Made Simple, Hove, Psychology Press Ltd Publishers.
Gray, Colin D., and Kinnear, Paul R., (1998)) SPSS for PC Made Simple, Hove, Psychology Press Ltd Publishers.

(4) **Do we really need three assessments ?**

If those who have come top in paper 1 are also those who have come top in paper 2 and the project, the question can be asked: why not just use one assessment and save everyone's time? Of course, if assessments are used as methods to make students work, then we need all three assessments!

To see to what extent the orders of merit in the three assessments are the same, we can use a statistical technique called ***correlation***. This calculates a *correlation coefficient* which is given the symbol: r.

If r = 1, then the orders of merit of two sets of data are identical. If r = -1, the orders of merit are inversely related - those top in one measurement are lowest in the other. If r = 0, then there is no relationship between the measurements.

There are two important points to note:

(a) The fact that we obtain a positive value for the correlation coefficient does ***NOT*** imply that there is any cause and effect relationship between the measurements.

(b) A positive correlation coefficient can arise by chance. However, a statistical package will give us an indication of this.

Again, although Excel can do the arithmetic, a statistical package like SPSS is much to be preferred. Pasting the numbers into SPSS, selecting statistics and correlation gives

Paper 1 and paper 2	r = 0.8
Paper 1 and project	r = 0.3
Paper 2 and project	r = 0.2

The correlation between papers 1 and 2 is very high indicating that the students who do well in one also do well in the other. We might deduce that the two papers are testing similar skills. allowing similar performances from students. However, the correlations with the project are very much lower, suggesting that the project assessment reflects a different set of skills.

In passing, we carry out correlations without standardising the marks in any way.

In SPSS, there are three ways to calculate the correlation coefficient. The handbook explains which should be used for which kind of data. In this case, Pearson correlation was used.

You need to be careful to use the correct correlation procedure. A simple guide is:

Pearson correlation	when data are interval (like test marks or scores)
Spearman correlation	when data are ordinal (when students are ranked in order)
Kendall's Tau-b correlation	when data are ordinal but there are few categories (like a Likert measurement)

Statistics Appendix 2

How to Handle Questionnaire Data

Looking at Percentages

To illustrate the way data can be handled when looking at attitudes and other characteristics of students, imagine physics students. The researcher is interested in what attracted these students towards physics and what is likely to retain them in physics in the future.

This illustration will use parts of a questionnaire which was actually used for this purpose and the data are based on the actual data obtained[10].

One question explored what were the likely factors which attracted the students into physics as a degree course. This question was given to 218 level 1 and level 2 students taking courses in physics and asked them to look back:

Which factor(s) influenced your choice of planned honours subject(s)?
Tick as many as you wish

- ☐ Enjoyment of subject
- ☐ Good grades at school in subject
- ☐ Your teacher at school
- ☐ Your parents
- ☐ Information from mass media
- ☐ Friends
- ☐ Likely career opportunities
- ☐ Demonstrations, exhibitions, festivals
- ☐ Any other factors (please list below)

....

It is possible and easy to add up all the 'ticks' for each choice and present these as percentages of the total number of students. [It is also possible to do the same for the men and women separately to explore whether there are gender differences in the pattern of ticks.] Because students can tick as many boxes as they like, the total of the percentages will not add to 100%

The overall results actually were:

87%	Enjoyment of subject	3%	Friends
74%	Good grades at school in subject	49%	Likely career opportunities
27%	Your teacher at school	9%	Demonstrations, exhibitions, festivals
9%	Your parents	7%	Any other factors (please list below)
4%	Information from mass media		....

There is no need to do more statistics on this. The pattern immediately tells us that, if we wish to attract more students into physics, then we can safely ignore most of the factors. *It tells us that the school experience is by far the dominant factor*. The quality of the school syllabus and the quality of the teachers are the critical factors. It is unlikely that we can influence the school physics curriculum although we might be able encourage able students to consider school teaching. However, any efforts in arranging special events, demonstrations, science centres, exhibitions are likely to have only marginal effects in attracting students towards physics and are unlikely to justify the effort, energy and expenditure. There is one area which is open to us - *the perception of career opportunities*. Other research in the same survey confirmed that this is an area where more information is needed by school pupils and students.

10 Skryabina, Elena, Attitudes Towards Physics, PhD Thesis, University of Glasgow, 1999. This is one of the most comprehensive studies in recent years on this topic, reflecting the Scottish school and university scene.

Grouping Data

Several questions explored what students thought of their experiences in their physics course. Two of these are now shown. The first uses the Semantic Differential format and illustrates the versatility of this approach to gain information very rapidly.

What are your opinions about University Physics?

Place a tick in one box between each phrase to show your opinions.

I feel I am coping well	☐	☐	☐	☐	☐	☐	I feel I am not coping well
I am not enjoying the subject	☐	☐	☐	☐	☐	☐	I am enjoying the subject
I have found the subject easy	☐	☐	☐	☐	☐	☐	I found the subject hard
I am growing intellectually	☐	☐	☐	☐	☐	☐	I am not growing intellectually
I am not obtaining new skills	☐	☐	☐	☐	☐	☐	I am obtaining new skills
I am enjoying practical work	☐	☐	☐	☐	☐	☐	I am not enjoying practical work
I am getting worse at the subject	☐	☐	☐	☐	☐	☐	I am getting better at the subject
It is definitely 'my' subject	☐	☐	☐	☐	☐	☐	I am wasting my time in this subject

The student choices were recorded and shown as percentages of the whole group. However, the researcher, in an attempt to make the data more accessible to the reader, presented it under three categories by grouping the six categories in pairs. This illustrates a number of important aspects of this kind of pedagogical research. If students are asked to respond on, say, a three point scale, they show some frustration. Experience shows that a 5 or 6 point scale is better. However, unless the number of students in the sample is huge, there may be small numbers in many of the categories. Grouping into, say, three categories makes the data more meaningful. However, grouping loses information and any statistical processing is better carried out on the original data wherever possible[11].

Here are the data grouped:

The data (as % of 165 level 1 students) were:

	%	%	%	
I feel I am coping well	36	47	17	I feel I am not coping well
I am not enjoying the subject	28	52	19	I am enjoying the subject
I have found the subject easy	6	53	41	I found the subject hard
I am growing intellectually	38	58	5	I am not growing intellectually
I am not obtaining new skills	41	46	13	I am obtaining new skills
I am enjoying practical work	28	43	27	I am not enjoying practical work
I am getting worse at the subject	8	58	32	I am getting better at the subject
It is definitely 'my' subject	24	59	17	I am wasting my time in this subject

Just by looking at the data, we can see that, despite finding the subject difficult, the students feel they are coping and developing. The one area of concern is the practical work. Subsequent enquiry identified the nature of the problem.

11 A Statistic like chi-square should be carried out on raw frequency data (not percentages). However, if the number of students selecting any category drops below about 10, then grouping is advised. Check a statistic textbook on this or seek advice from a statistician in your university.

Comparing Attitudes Between Groups

In her work, the researcher looked extensively at school pupil attitudes as it became very clear that attitudes towards physics are formed early and largely control choices of future studies in physics. To illustrate how attitude data can be handled, here are some data she obtained for school pupils at three stages (ages approximately 14-15, 15-16, 16-17). Pupils were asked about physics lessons:

I like physics lessons	☐	☐	☐	☐	☐	☐	I hate physics lessons
boring lessons	☐	☐	☐	☐	☐	☐	interesting lessons
easy lessons	☐	☐	☐	☐	☐	☐	complicated lessons
useless lessons	☐	☐	☐	☐	☐	☐	important lessons
enjoying lessons	☐	☐	☐	☐	☐	☐	boring lessons

Age 14-15 - 103 pupils   Age 15-16 = 152 pupils   Age 16-17 = 96 pupils

Again, she grouped the data in three categories (as % of samples)

	Age 14+	54	36	10	
I like physics lessons	Age 15+	55	36	8	I hate physics lessons
	Age 16+	38	55	7	
	Age 14+	11	47	42	
boring lessons	Age 15+	7	43	50	interesting lessons
	Age 16+	14	57	29	
	Age 14+	21	57	22	
easy lessons	Age 15+	9	67	25	complicated lessons
	Age 16+	13	55	32	
	Age 14+	3	34	63	
useless lessons	Age 15+	0	26	74	important lessons
	Age 16+	2	59	39	
	Age 14+	47	40	13	
enjoying lessons	Age 15+	54	36	8	boring lessons
	Age 16+	31	53	16	

Looking at the data, are the differences between the year groups significant or is it likely that they arose by chance. The statistic χ^2 is used to answer this question. After calculating the value, a table of chi-square values is consulted to see if the result is significant.

The results for the first question are:

Are the 15+ group different from the 14+ group ?	$\chi^2 = 0.6$	Not significant	Groups are similar
Are the 16+ group different from the 15+ group ?	$\chi^2 = 14.8$	$p < 0.01$	Groups are different
Are the 16+ group different from the 14+ group ?	$\chi^2 = 15.0$	$p < 0.01$	Groups are different

Warning: The control group can be defined in different ways when using chi-square. When using a computer package, it is not always easy to know what is happening. Check with a statistician or check a calculation by hand.

The Higher Education Academy
UK Physical Sciences Centre

...enhancing the student experience in chemistry, physics, astronomy and forensic science within the university sector.

Physical Sciences Practice Guides are designed to provide practical advice and guidance on issues and topics related to teaching and learning in the physical sciences. Each guide focuses on a particular aspect of higher education and is written by an academic experienced in that field.

This booklet is written for the many lecturers in the physical sciences who have a concern for and interest in the teaching of their disciplines. It is a challenging task to share something of what we know with future generations to enable them to discover the excitement of exploring the physical world.

In the annotated bibliography [available at <www.heacademy.ac.uk/physsci/publications/practiceguides>], an attempt has been made to summarise the main findings from pedagogical research in the physical sciences. This text seeks to introduce the ways in which our understandings of teaching and learning can be explored.

It is hoped that readers will find this booklet useful as a guide and pattern as they begin their own pedagogical research.

Norman Reid is Head of the Centre for Science Education, University of Glasgow.