EUROPEAN COMMISSION

Targeted Socio-Economic Research Programme

Project *PL 95-2005*

LABWORK IN SCIENCE EDUCATION

* WORKING PAPER 4 *

SURVEY 2: STUDENTS' IMAGES OF SCIENCE AS THEY RELATE TO LABWORK LEARNING

John Leach, Robin Millar, Jim Ryder, Marie-Geneviève Séré
Dorte Hammelev, Hans Niedderer, Vasilis Tselfes

1998

Contact Details: Dr. John Leach
LIS, CSSME,
The University of Leeds
Leeds LS2 9JT
UK

Tel. +44 113 233 4679
Fax. + 44 113 233 4683
Email:
J.T.Leach@education.leeds.ac.u
k

Improving Science Education: Issues and Research on Innovative Emprical and Computer-Based Approaches to Labwork in Europe

Short Title : Labwork in Science Education

Financed by DGXII of the European Commission between February 1996 and April 1998.

The Partnership:

France - DidaSco Université Paris XI and INRP
Marie-Geneviève Séré (Co-ordinator and Group Leader), Daniel Beaufils, Michel Beney, Alain Guillon, Didier Pol, Nahim Salamé, Jean Winther

Denmark - FIFU, Regional Centre of research and promotion of Further education
Albert C. Paulsen (Group-leader, Royal Danish School of Educational Studies), Dorte Hammelev (Roskilde University), Helge Kudahl (FIFU)

France - COAST - GRIC Université Lyon 2
Andrée Tiberghien (Group Leader), Karine Beçu-Robinault, Christian Buty, Jean-François Le Maréchal, Laurent Veillard

Great Britain - LIS, The University of Leeds; The University of York; King's College, London
John Leach (Group Leader, The University of Leeds), the late Rosalind Driver (King's College), Jenny Lewis (The University of Leeds), Robin Millar (University of York), Jim Ryder (The University of Leeds)

Germany - Institut für Didaktik der Physik, University of Bremen; University of Dortmund
Hans Niedderer (Group Leader), Stefan non Aufschnaiter, Hans Fischer, Kerstin Haller, Lorenz Hucke, Florian Sander, Horst Schecker, Manuela Welzel

Greece - TESME, Aristotle University of Thessaloniki
Dimitris Psillos (Group Leader), A Basrbas, Garabet, Garo Bisdikian, Dimitris Evangelinos, Petros Kariotoglou, Koumaras, Anastasios Moholidis, Vasilis Tselfes

Italy - University of Rome 'La Sapienza'; University of Rome 3
Matilde Vicentini (Group Leader), Milena Bandiera, Francesco Dupré, Carlo Tarsitani, Eugenio Torracca

Outcomes

A list of the full set of Working Papers from the project can be found at the end of this document. Further results from this work can be found on the Internet via the CORDIS site of the European Commission : http://www.cordis.lu/
The abstract of the project provided on this site is given on the next page.

ISBN: 0-904-42192-9

ABSTRACT: *'Labwork in Science Education '*

This project stems from a concern to recognise science education as an important component of a general education, not only for future scientists and engineers, but also for any future citizen in a European society which is increasingly dependent upon science and technology.

Research has focused upon the role of laboratory work ('labwork') in science teaching at the levels of **upper secondary school and the first two years of undergraduate study**, in physics, chemistry and biology. Various forms of labwork have been identified and investigated, including 'typical' activities in which pairs of students work on activities following precise instructions, open-ended project work in which students design and carry out empirical investigations, and the use of modern technologies for modelling, simulating and data processing.

The main objectives of the project were to clarify and differentiate learning objectives for labwork, and to conduct investigations yielding information that might be used in the design of labwork approaches that are as effective as possible in promoting student learning.

A survey was conducted to allow for better description of existing labwork practices in the countries involved. There are great variations from country to country in the time devoted to labwork, the assessment of students' performance in labwork and the equipment available. However, the forms of labwork activity used between countries are remarkably similar. In each country, the most frequent activity involves students following precise instructions in pairs or threes. A document has been produced describing the place of labwork in science education in each country.

A second survey was conducted to study the learning objectives attributed to labwork by teachers. There are some differences between countries in terms of the relative importance given to the teaching of laboratory skills. Motivation for science learning is not attributed particularly high status as an objective for labwork learning. In each country, the main goal for labwork teaching in the view of teachers surveyed concerns enabling students to form links 'between theory and practice'.

A third piece of survey work was conducted to investigate the images of science drawn upon by students during labwork, and the image of science conveyed to students by teachers during labwork. These surveys were based upon the hypothesis that epistemological and sociological ideas about science are prominent during labwork.

22 case studies were carried out in order to clarify the variety of knowledge, attitudes and competencies that can be promoted through labwork. The case studies focused upon both empirical labwork and labwork involving computer modelling and simulation. The work has resulted in an analysis of the **effectiveness of labwork**, leading to recommendations about policy. It is hoped that teachers and policy makers with responsibilities in science education generally, and labwork in particular, will find these useful in informing future practice with respect to possible objectives for

labwork, links between objectives, methods of organisation of labwork and ways of observing and evaluating the effectiveness of labwork in promoting student learning.

Contents

1	**Images of science and the teaching laboratory**	**1**
2	**Aspects of students' images of science that influence learning during labwork**	**3**
	Hypotheses about students' images of data and measurement	*4*
	Hypotheses about students' images of the nature of investigation	*6*
	Hypotheses about students' images of the nature of theory	*7*
	Hypotheses about students' images of the nature of explanation	*7*
	Hypotheses about students' images of the nature of public scientific knowledge	*8*
3	**Investigating the images of science that might influence students' learning during labwork: methodological issues**	**8**
	The issue of language	*8*
	Methodological perspectives in the literature	*9*
	Contextualisation of the survey in student labwork	*12*
	Approach used in the design of diagnostic questions and their analysis	*13*
4	**The five probes: diagnostic questions designed to investigate the images of science that might affect students' learning during labwork**	**15**
	Differences in values from measurement	*16*
	Theory and data in scientists' work	*17*
	The nature of scientific results	*18*
	Surprising results	*19*
	Interpreting data	*21*
	General methodological issues	*23*
5	**Details of sample**	**26**
6	**Analysis and findings**	**28**
	Differences in Values from Measurement	
	Sample for this probe, and reporting categories	*28*
	Tabulation of data	*29*
	Discussion of results	*33*
	Theory and Data in Scientists' Work	***34***
	Tabulation of data	*35*
	Discussion of results	*37*
	The Nature of Scientific Results	***38***
	Tabulation of data	*38*
	Discussion of results	*42*

Surprising Results ***42***
Tabulation of data *42*
Discussion of results *46*

Interpreting Data ***47***
Tabulation of data *47*
Discussion of results *53*

The role of context ***53***

7 Discussion: the significance of findings from the study of labwork **56**
Summary of findings ***56***
Recommendations for labwork teaching arising from the study ***60***

Acknowledgement **61**

References **62**

Appendix 1: The full text of probes **64**

Labwork in Science Education, some policy implications: A summary **78**

Working Papers of the project 'Labwork in Science Education' **82**

Survey 2: Students' 'images of science' as they relate to labwork learning

John Leach, Robin Millar, Jim Ryder, Marie-Geneviève Séré, Dorte Hammelev, Hans Niedderer, Vasilis Tselfes[1]

ABSTRACT
When students carry out labwork they do so based on assumptions about the purposes of the task in hand, the nature of the data that might be collected and their relationship to knowledge, the nature of explanation and investigation and so on. In the project *Labwork in Science Education* we have attempted to investigate the images of science that European science students in upper secondary schools and universities might draw upon during labwork, and the impact that these images of science might have upon students' learning.

In the first part of this paper, possible links between students' images of science and their performance during various sorts of labwork are considered. Then, a number of diagnostic questions used to investigate students' images of science are presented, consideration being given to the methodological advantages and disadvantages of various approaches. Details of data collection and analysis, and findings, are then presented. Finally, the significance of findings from the survey for teaching and learning through labwork are discussed.

1 Images of science and the teaching laboratory

This paper describes an investigation of the images of science found amongst science students in upper secondary schools and at the beginning of university, that might influence their learning during labwork. The phrase 'images of science' is used as a shorthand to refer to representations of the epistemology and sociology of science used by individuals in specific contexts for specific purposes. We have assumed that individuals hold, and draw upon, a variety of representations of the epistemology and sociology of science in different contexts (Driver et al., 1996). The particular focus of this study is upon the images of science drawn upon and used by students during their labwork learning, and no attempt has been made to elicit images of science that might be drawn upon by students in contexts other than labwork.

What have 'images of science' got to do with labwork? To answer this question, we need to ask ourselves about the purposes of labwork itself. During the project *Labwork in Science Education* a conceptual exercise was undertaken to identify the different purposes of labwork (Millar et al., 1998). The purposes identified can be grouped into three broad areas:

---

[1] The hypotheses, research questions and probes described in this paper were produced during an extensive period of collaboration between the participating countries. The following researchers were involved in the elaboration of research questions and hypotheses:

Milena Bandiera, Rosalind Driver, Dimitris Evangelinos, Dorte Hammelev, Helge Kuhdal, John Leach, Jenny Lewis, Jean-François le Marechal, Robin Millar, Hans Niedderer, Jonathan Osborne, Albert Christian Paulsen, Dimitris Psillos, Jim Ryder, Marie-Geneviève Séré, Carlo Tarsitani, Andrée Tiberghien, Eugenio Torracca, Vasilis Tselfes, Matilde Vicentini, Jean Winther.

During this collaboration, many differences of opinion were expressed. The authors therefore take full responsibility for the opinions expressed in this paper, while acknowledging the contributions of the researchers listed above. The feedback and comments of other members of the LSE project is also greatfully acknowledged.

- developing students' knowledge of the behaviour of the natural world, helping them to make links between the world of natural phenomena and the world of theoretical descriptions and explanations of phenomena, and thereby developing their understanding of scientific concepts;
- developing students' understanding of how scientists undertake empirical investigations to address a question or problem of interest;
- developing students' ability to use standard laboratory instruments and procedures in undertaking investigations

The first of these areas is mainly concerned with teaching scientific *content* (i.e. the laws, theories, concepts etc. that constitute scientific knowledge). By contrast, the main concern of the second two areas relates more to teaching about the *methods* used by scientists in empirical work. A further aim of some labwork relates to teaching about the social and institutional processes followed in scientific communities (e.g. Ryder et al., in press).

Labwork with each of these aims involves students in drawing upon understandings of the nature of empirical data, the nature of scientific knowledge claims, the ways in which knowledge claims and data are related, the purposes of using techniques, procedures and instruments, and so on. Many students in teaching laboratories often work with knowledge claims already agreed as reliable within the scientific community. For example, they may be involved in work to illustrate accepted theories or to apply accepted theory in specific contexts. Their ideas about how that knowledge came to be viewed as reliable may well influence their labwork. For all these reasons, participation in labwork involves students in drawing upon epistemological understanding.

There is a good deal of evidence that the images of science drawn upon by many students during labwork constrain performance. For example, Séré et al. (1993) have illustrated how university students' understanding of the nature of data result in them taking inappropriate actions during labwork that involves measuring physical quantities. Ryder et al. (1997) have shown how university students working on open-ended investigations sometimes draw upon understandings of the relationships between data and knowledge claims that result in inappropriate actions being taken. Guillon and Séré (1998) show that students experience similar difficulties in more closed labwork (Case Study FD3, reported in Psillos et al., 1998).

A number of studies of students' images of science have been reported in the literature, and a number of these are referred to later in this paper. However, few of these relate either to students of upper secondary and university age or to the images of science that students might draw upon during labwork. The study reported in this paper therefore had two major purposes:

- To review existing research on students' images of science, and to consider the nature of labwork, in order to propose hypotheses linking students' images of science with their learning during labwork.

- To generate preliminary baseline data about upper secondary and university science students' images of science, in areas hypothesised as being relevant to labwork learning.

It is hoped that the baseline information about students' images of science generated in this study will be used in a number of ways. It was suggested earlier that the purpose of some labwork teaching is to teach students about specific aspects of the nature and practice of science. The baseline information provided by this study will hopefully be useful to teachers in identifying the likely starting points of students, thereby allowing them to focus teaching more effectively upon students' needs. In the case of labwork aiming to teach science content, it is intended that teachers will also draw upon this baseline information to help them to identify some of the difficulties likely to be experienced by students as a result of the use of inappropriate images of science. Many European science curricula at the upper secondary and university levels specify learning goals for labwork. The baseline data generated in this study will hopefully be useful in refining such curriculum goals, as specific areas that students find difficult are highlighted.

No attempt has been made in this survey to analyse national differences in images of science. This is due to the inherent difficulties in constructing comparable samples that are also nationally representative. For example, students in Great Britain follow a specialised curriculum from the age of 16, whereas science students in other countries follow a broad curriculum until well into their university studies.

The extent to which students' responses to survey questions provide insights into the images of science used by students actually engaged in labwork is an open question. During the project *Labwork in Science Education*, a number of case studies are being conducted, to examine the images of science drawn upon by students in actual teaching contexts (see Psillos et al., *Working Paper 7*). The insights provided by some of these case studies do suggest that many students use similar images of science in response to survey questions, and during labwork (e.g. Case Study GB1, Leach; GB3, Ryder; GR2, Kariotoglou).

2 Aspects of students' images of science that influence learning during labwork

At the beginning of this paper, a case was made that when students engage in labwork they may draw upon understandings of the nature of data, the nature of knowledge claims, links between data and knowledge claims or the warranting of public knowledge as reliable. In this section, hypotheses are presented linking students' images of science and their learning in labwork, together with research questions for the study.

The hypotheses relate to 5 broad aspects of students' images of science, namely the nature of data and measurement in empirical work, the nature of investigation in science, the nature of theory in science, the nature of explanation in science and the nature of reliable public scientific knowledge. In each case, details are given linking students' images of science and their learning during labwork. There are 9 hypotheses in total, and possible research questions are presented for each one.

Hypotheses about students' images of data and measurement

HYPOTHESIS 1 (H1)
Many students consider that, with good enough apparatus and enough care, it is possible to make a perfect measurement of a quantity. That is, they assume that measurement can be perfectly *accurate*.
Others consider that any measurement is subject to some uncertainty, and so obtaining accurate values is problematic.
(e.g. Séré et al., 1993; Lubben and Millar, 1996)

Why it is relevant to labwork
During labwork, students may have to make decisions about the amount of data that has to be collected and the conclusions that can be drawn from given data sets. The decisions that they make about this aspect of data collection will be influenced by their view of the nature of measurement. For example, students who see measurements as 'perfect' may join each individual data point on a graph rather than plotting lines or curves of best fit. Similarly, in deciding upon a value from a set of measurements they are likely to select the mode, reasoning that the most frequently recorded value must be the 'true' value. Others recognise that all measurements add information, and can therefore be treated as a set using statistical techniques. Amongst these students, some assume that statistical calculations will yield information about the *accuracy* of measurement, whereas others recognise that such calculations only yield information about *precision*.

Research Questions
RQ1: Do students see measured data as a 'perfect' copy of reality, or do they view measured data as being subject to some uncertainty? What do they see as the sources of uncertainty in measured data? How do they overcome these uncertainties and select a value? Do they recognise the difference between *accuracy* and *precision*?

HYPOTHESIS 2 (H2)
Some students do not recognise the kinds of empirical evidence on which scientific knowledge claims are based. In the case of measured data, they think that it is only possible to judge the quality of a measurement from a knowledge of the 'true' value, given by an authority source. That is, they do not recognise that decisions about *precision* can be made from sets of measurements.
Other students think it is possible to judge the 'quality' of measured data from a set of repeated measurements. That is, they reason that data sets can be evaluated in their own terms to make decisions about accuracy and precision.
(e.g. Séré et al., 1993; Lubben and Millar, 1996)

Why it is relevant to labwork
Claims about the values of measurements, whether made by students in labwork classes or authority sources in data books, are based on empirical measurements. If students do not recognise the relationships between knowledge claims and empirical evidence, they are likely to approach data collection and interpretation during labwork differently from students who believe that the quality of a measurement can be judged from a set of repeated measurements. Some students assume that mean values from

sets of repeated measurements give an indication of the *accuracy* of a measurement, whereas others recognise that statistical processing of a data set only gives an indication of the *precision* of a measurement.

Research Questions
RQ2: Do students believe that the only way to judge the quality of a measurement is from a known ‘true’ result, or do they believe that the quality of a measurement can be judged from a set of repeated measurements? If so, do they distinguish the *accuracy* and *precision* of measured values?

HYPOTHESIS 3 (H3)
Many students see data reduction and presentation as a process of *summarising* data and see procedures like joining data points on a graph, drawing a ‘best fit’ straight lines, or drawing smooth curves as *routine heuristics* - that is, they see the process as independent of theory. They believe that there are standard techniques for arriving at ‘perfect’ descriptions of data.
Others see such procedures as a process of proposing *tentative hypotheses about a relation between variables.* That is, they believe that experimenters (and computers) make decisions during data reduction and presentation according to existing models.
(e.g. Séré et al., 1993; Lubben and Millar, 1996)

Why it is relevant to labwork
The types of conclusions drawn by students during labwork will be influenced by the students’ views of the nature of data handling. For example, once data points are plotted on to axes of a graph, hypotheses have to be proposed about possible relationships between the points. Students who see each data point as a ‘perfect’ value may well join each point. They may reject lines of best fit that do not pass through any data points. Other students who see procedures such as linear regression as routine heuristics may well apply one procedure without considering whether it is valid to do so, or whether other procedures may be more valid. In addition, they may well view scientific knowledge as akin to algorithms, leading to unique results. Similarly, data sets from laboratory work may be treated as the unique products of algorithms.

Research Questions
RQ3: When working with data sets, do students see procedures like joining data points with lines of ‘best fit’ or smooth curves as routine heuristics, or alternatively as a process of proposing tentative hypotheses?

HYPOTHESIS 4 (H4)
Some students think that the logic of proof and falsification is symmetrical: data that logically support a law ‘prove’ the law, in the same way that data that do not support a law logically falsify it.
(e.g. Kuhn et al., 1988; Driver et al., 1996)

Why it is relevant to labwork
During labwork, and particularly open-ended labwork, students’ approaches to data collection and data processing may be influenced by their beliefs about the logic of proof and falsification.

Research Questions:
RQ4: Do students recognise the logical distinction between proof and falsification when handling empirical data?

Hypotheses about students' images of the nature of investigation

HYPOTHESIS 5 (H5)
Some students think that most/all questions about natural phenomena are answerable by collecting observational data and looking for correlations. Explanatory theories (models) 'emerge' from this data in a logical way: there is only one possible interpretation.
Other students think that prior models (theories, hypotheses) influence decisions about what data to collect and how it is interpreted, and that observation and measurement are intended to test these models. Again, the testing is based on logic: only one interpretation is possible.
Others think that a data base is first collected on the basis of embryonic theories and hypotheses - more robust models are then proposed as conjecture to account for existing, and anticipated data. Then predictions derived from these may be tested by planned observations or experiments, but more than one interpretation is possible due to the conjectural nature of theory.
(e.g. Driver et al., 1996; Larochelle and Désautels, 1991; Aikenhead et al., 1987; Niedderer et al., 1992)

Why it is relevant to labwork
Students' approaches to data collection and data interpretation during labwork will be influenced by their views of the place and nature of theory in empirical investigation. They may not recognise the interplay between data and theory in the process of investigation, and as a result they may not accept that it is legitimate to develop the design of an experiment in the light of data already collected.

Research Questions
RQ5: Do students think that scientific theories 'emerge' from data, or do they think of scientific theories and data as being related in a more complex way? If so, how do they think that scientific theories and data are related? In particular, do they think that a given experiment is open to more than one interpretation?

HYPOTHESIS 6 (H6)
Many students see practical activities in the teaching lab as exercises to reproduce well-known results, or to illustrate important theories/models, no matter how the task is actually presented by the teacher! They do not recognise the labwork as an exercise in 'finding out'.
Other students recognise that some labwork activities have an investigative component: they involve 'finding out'. Amongst these students, some assume that knowledge claims can be 'proved' or 'disproved' by a single planned intervention, whereas others assume that the process of investigation involves a sequence of interventions which may be modified in the light of experience.
(e.g. Driver et al., 1996; Lubben and Millar, 1996)

Why it is relevant to labwork
The actions of students during experimental design, data collection and data interpretation in labwork will be greatly influenced by their views of the purpose of the labwork task.

Research Questions:
RQ6: Do students recognise the purpose of particular labwork tasks as involving ‘finding out’?

Hypotheses about students’ images of the nature of theory

HYPOTHESIS 7 (H7)
Some students believe that scientific theories are really descriptions of natural phenomena: there is a one-to-one correspondence between theory and ‘reality’. Such students believe that it is a straightforward empirical process to show that scientific theories are ‘true’.
Others believe that theories are model-like, and do not simply describe reality. However, such students still believe that it is a straightforward empirical process to show that scientific theories are ‘true’.
Others believe that theories are model-like, and this means that it is NOT a straightforward process to show that a scientific theory is ‘true’.
(e.g. Larochelle and Désautels, 1991; Aikenhead et al., 1987; Driver et al., 1996; Niedderer et al., 1992)

Why it is relevant to labwork
Students who do not recognise the conjectural nature of models in science may frame their data interpretation during labwork in terms of observational features of the phenomenon, rather than theoretical entities.

Research Questions:
RQ7: Do students think that scientific theories are conjectural and model like in nature, or do they think that theories are essentially descriptions of phenomena in different terms?

Hypotheses about students’ images of the nature of explanation

HYPOTHESIS 8 (H8)
Some students do not recognise the different levels, types and purposes of explanation that are used in science. [Examples: teleological, causal, descriptive, model-based..]. (e.g. Tamir and Zohar, 1991; Leach et al., 1996)

Why it is related to labwork
The types of conclusions that are drawn by students during labwork, and particularly open-ended labwork, will be influenced by the type of explanation that the student thinks is most appropriate.

Research Questions:

RQ8: Are students able to distinguish between teleological, descriptive and model-based explanations of natural phenomena?

Hypotheses about students' images of the nature of public scientific knowledge

Hypothesis 9 (H9)
Some students think that all the knowledge claims made by science are of the same status. They do not recognise the role of the scientific community in the validation of public knowledge.
Others recognise that some knowledge claims are widely accepted within the scientific community, whereas others are still the subject of investigation and debate.
(e.g. Driver et al., 1996)

Why it is relevant to labwork
If students believe that all knowledge claims in science are of equal status, this will affect their actions during labwork, and particularly open-ended labwork. If data collected during labwork are not consistent with canonical science, a number of options are open to the investigator: the option that is pursued will depend upon the investigators' beliefs about the status of the scientific knowledge in question.

Research Questions:
RQ9: Do students recognise that different courses of action are appropriate in scientific investigations depending on the status of the scientific knowledge claim under investigation?

These hypotheses and the associated research questions were drawn upon in designing and selecting diagnostic questions for the survey.

3 Investigating the images of science that might influence students' learning during labwork: methodological issues

The issue of language

Collecting and analysing data from students in six countries with different national languages raised specific methodological issues. In practice any research instruments produced have to be translated, as do any data written in students' own words. The process of making translations whilst minimising changes of meanings is problematic to say the least. Research instruments could be administered as individual or group interviews or as paper and pencil exercises. If students respond in their own words, then the resulting data set will contain responses in several languages. How might coding schemes be generated that reflect possible national differences? One approach is to translate all data into a language understood by all the researchers (in this case, English). However, the problems of producing translations that reflect the nuances of students' responses cannot be overemphasised. Another approach is to use closed-format responses, students recording their opinions by making ticks against pre-defined statements. This approach also has its problems, in that the wording of closed-response boxes has to be interpreted similarly no matter what language the statement is written in.

It was decided to generate a paper and pencil questionnaire to investigate the images of science that might influence students' learning during labwork. Although an interview study would have generated data better able to yield rich insights into students' reasoning, it was not deemed feasible to conduct such a study within the resources of the project. It was decided that the response mode for written questions should include as many closed-response items as possible, to minimise the amount of translation required and thereby reduce the potential for changing meanings through translation. The method used for generating closed response options is described later in this paper.

Closely related to the issue of language is that of the culture of science teaching and learning in the participating countries. Writing diagnostic questions that the participating researchers felt would have credibility amongst students at the upper secondary and university levels in each country proved highly problematic. Indeed, reservations about the face validity of a number of the questions remain. Some of these reservations are discussed later in this paper.

Methodological perspectives in the literature

There are a number of reports of students' images of science in the literature, many of which are based upon the use of written survey instruments and closed-response questions. One possibility was to take such an instrument and use it within the project in its original form, or with minor modifications. Another possibility was to draw upon methodological approaches used by others in the writing of a new instrument. The methodological approaches of some instruments reported in the literature were reviewed, with a view to ascertaining whether the approaches or instruments might be drawn upon within the project.

The methodological underpinnings of survey questions reported in the literature are summarised in the following paragraphs.

How are views of the nature of science used in framing questions to probe images of science?
The questions in some previously published surveys have been framed around particular positions in the philosophy of science. Consider the following statements which were to be rated 'true' or 'false' by science teachers:

- The results that pupils get from their experiments are as valid as anybody else's
- Science facts are what scientists agree that they are
- Scientific theories describe a real external world which is independent of human perception

Statements such as these were used to position individual teachers images of science on an axis labelled 'Relativism/Positivism'. We might describe this methodological approach as *nomothetic* in that questions are written around a normative 'map' of the nature of science. [This activity (Nott and Wellington, 1993) was designed as a training activity for teachers rather than a research questionnaire.]

An alternative approach involves writing survey questions that do not map directly on to any one philosophical position. Knowledge about the nature of science is drawn upon to identify broad areas of interest in people's images of science. Consider the following questions (Ryder et al., in press):

- How can good scientific work be distinguished from bad scientific work?
- Why do you think that some scientific work stands the test of time whilst other scientific work is forgotten?
- How are conflicts of ideas resolved in the scientific community?

This set of questions was designed to probe undergraduate students' images of how public scientific knowledge is agreed. Perspectives and issues in the history, philosophy and sociology of science were used to identify broad areas of students' reasoning for investigation, and open questions were written around which students and researchers could have extended conversations about these issues. It is possible for students to answer each question from a range of perspectives.

The study reported in this paper focuses upon the images of science that influence students' learning during labwork. In practice, very little published work was identified that made specific links between students' images of science and their learning during labwork. Draft diagnostic questions were written around the hypotheses and research questions, in some cases drawing heavily upon previously published diagnostic questions.

Generalised questions, or questions in context?

The examples of diagnostic questions referred to above are based upon 'science' in the abstract, with no clues about context. The obvious problem with decontextualised questions is that students are likely to have a variety of contexts in mind when answering the question, resulting in different answers being given. For example, a student thinking about their own labwork might say that good and bad scientific work can be distinguished by looking at the level of detail used in the written report, or the extent to which the empirical results generated are in line with accepted theory. By contrast, a student thinking about the work of a professional scientist might place more emphasis on the contribution of a particular piece of work to a broad field of enquiry. If no information about context is presented or elicited, at the stage of analysis researchers have no way of knowing which contexts students had in mind when answering the questions. The less information that is given about context, the less confident researchers can be as to students' interpretations of questions, and the intended meanings of students' responses. Furthermore, translating some of the terms often used in decontextualised questions (theory, fact, valid etc.) between 6 languages is highly problematic, compared to producing a description of a particular context in several languages.

An alternative approach to this involves asking questions in a specific context. Consider the following examples:

- Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not to give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug? (Miller, 1994)

- Community or government agencies should tell scientists what to investigate; otherwise scientists will investigate what is of interest only to them.

  Your position basically
  Community or government agencies should tell scientists what to investigate:
  A so that the scientists' work can help improve society.
  B only for important public problems; otherwise scientists should decide what to investigate
  C All parties should have an equal say. Government agencies and scientists together should decide what needs to be studied, even though scientists are usually informed about society's needs.
  (other choices given....)
  (Aikenhead, Fleming and Ryan, 1987)

In these cases, individuals' images of science are elicited in particular contexts. The statements A, B, C etc. in the work of Aikenhead et al. were generated from pilot work with students.

This sort of approach allows researchers to be more clear about the contexts that students had in mind when answering questions. However, findings elicited in a tightly-defined context cannot be applied to other contexts unproblematically. In addition, tensions are inevitable between presenting features of a context in a manner that is understood, and not misrepresenting the context by over-simplification or caricaturisation. In addition, it cannot be assumed that students' understandings of the context are the same as the meanings that the researchers intended to communicate.

Another approach to probing students' images of science that has been reported is to pose questions about particular investigations that the students are carrying out (e.g. Carey et al., 1989). For our purposes, this approach has the distinct advantage of eliciting students' images of science in the context of their learning during labwork. Data collection and analysis across a variety of different types of labwork is taking place through case studies in the project *Labwork in Science Education* (see Psillos et al., *Working Paper 7*). In addition, it is intended that this survey will provide more general baseline information about students' images of science in the appropriate age range.

Methodological decisions taken in the light of approaches reported in the literature
It was decided to write a number of diagnostic questions in which a context is specified, to avoid the problems of interpretation referred to above. In addition, however, some decontextualised questions were also written, allowing individual students' responses to be compared on contextualised and decontextualised questions.

Decontextualised questions are generally much shorter and quicker to answer, and so if students' responses to them correlate well with their responses to contextualised questions, they might be a more efficient tool for survey work.

Contextualisation of the survey in student labwork

The aim of the survey was to investigate the images of science that might influence students' learning during labwork. In practice, previous research on students' images of science has addressed several areas including the nature of the questions addressed by science, the nature of empirical data in science, the nature of scientific explanation and the nature of reliable public knowledge (reviewed by Driver et al., 1996). For our purposes, however, it was necessary to generate more specific links between students' images of science and their learning during labwork and for this reason the research hypotheses and questions were written.

Four types of context were identified for investigating the images of science that might influence students' learning during labwork:

- *Responses to laboratory situations*
  Nott and Wellington (1997) describe the use of 'critical incidents' in the teaching laboratory to probe teachers' images of science. A typical situation is presented to teachers, such as data from a standard school experiment not agreeing with accepted theory. Teachers are then asked what they would do in this situation, and the reasons for this. It was felt that a similar approach could be used in this study to investigate the images of science drawn upon by students in specific labwork situations, particularly as they relate to relationships between knowledge claims and data. The 'critical incidents' used might be located in the teaching laboratory or the research laboratory.

- *Questions about the reliability of public knowledge statements*
  In order to investigate views on the warranting of public knowledge statements as reliable, it was envisaged that students could be presented with examples of public knowledge statements, and be asked why the statement is believed to be reliable. A similar approach has been used by Driver et al. (1996).

- *Questions comparing school science and professional science*
  In order to gain insights into students' views of the relative importance of empirical investigation and opinion within the scientific community in warranting knowledge claims as reliable, it was envisaged that questions could be designed which asked students to explain their views about differences between empirical work in the teaching laboratory, and that carried out by professional scientists.

- *Context-free questions*
  As stated earlier it was decided to use some context free questions, in an attempt to elicit information about students' general epistemological orientations.

Approach used in the design of diagnostic questions and their analysis

Diagnostic questions were designed through an iterative process involving a large number of researchers in each participating country. At the beginning of the project, it was decided that each national group should highlight the aspects of students' images of science that they thought might relate to the students' learning during labwork, and write some draft diagnostic questions that might be used to probe students' images of science in these areas. Papers were produced by each national group and circulated. A commentary on areas of similarity and difference amongst the papers was also produced and circulated.

A meeting of researchers was then convened, the aim of which was to identify aspects of students' images of science to be investigated, and to identify diagnostic questions to be developed and piloted for this purpose.

During this meeting, 41 draft diagnostic questions were discussed. In each case, one of the following decisions was made about future work on the question:

- *a* *This issue investigated by this question is interesting and relevant. We will develop the existing question for piloting;*
- *b* *This issue is interesting and relevant, but we are not sure how to investigate it. We will try and work up a question to pilot. If we are successful, we will include the question in the survey. If we are unsuccessful, this area will not be included in the survey;*
- *c* *The issue is interesting and relevant, but we are not going to pursue it in the survey for pragmatic reasons;*
- *d* *This issue is either uninteresting or irrelevant, so we will not pursue it any further.*

By the end of the meeting, 4 questions had been coded 'a', 7 had been coded 'a/b' and 11 were coded 'b'. There was great overlap in the focus of these questions. Responsibility for coordinating further development of these questions was given to the GB group, in some cases in collaboration with colleagues in other groups. The aim was to produce a draft survey instrument of a manageable size for piloting in 3 months, with a view to presenting findings of the pilot in another month.

In order to produce a manageable number of diagnostic questions for piloting, the GB group sorted the 22 draft diagnostic questions identified for further work into a number of clusters, each addressing broadly similar aspects of students' images of science. Questions within these clusters were then drawn upon to produce a much smaller number of draft diagnostic questions for piloting. In addition, attempts to develop some clusters of questions were abandoned, mainly because the GB group did not feel, in hindsight, that they could be developed to elicit useful and interesting information about students' images of science as they might relate to learning during labwork.

This work took place during the summer of 1996, and as a result only a small number of researchers outside of GB were available to comment on work in progress within the deadlines imposed by the work plan for the project.

As a result of this process, 5 Draft Diagnostic Questions were written and presented to researchers in the project. A number of issues and problems were raised at this stage relating to the coverage of the survey, and the formulation of the Draft Diagnostic Questions produced by the GB group. In order to address some of these problems, it was decided to begin a process of writing hypotheses and research questions (the products of which were reported earlier in this paper) and to make further modifications to the Draft Diagnostic Questions before undertaking piloting. In addition, all researchers in the project were invited to make a critique of the Draft Diagnostic Questions.

Following this development work, 5 new Draft Diagnostic Questions were developed and piloted in a number of countries. The aim of this pilot study was to assess the face validity and communicability of the questions with students, and to generate data from questions with an open-format response mode to be used in the generation of closed-response items.

The pilot sample size used is shown in Table 1:

TABLE 1: PILOT SAMPLE DETAILS

COUNTRY	*UPPER SECONDARY STUDENTS*	*UNIVERSITY STUDENTS*
Denmark	11	0
France	75	125
Germany	0	49
Great Britain	36	0
Greece	0	13
TOTAL	115	187

In addition, a number of other diagnostic questions were piloted in France, Germany, Greece and Italy.

Analysis of this pilot data set was carried out independently in each country. One approach to data analysis is to code responses in terms of a normative view of the nature of science (Nott and Wellington, 1993). An alternative approach involves identifying coding categories and reporting units in terms of actual responses to survey questions (an *ideographic* approach). The pilot sample was analysed ideographically, broad categories in students' responses to open-ended questions being identified. This was felt to be the best way of probing the images of science likely to be drawn upon by students in action situations.

The analysis carried out in each country on the 5 draft diagnostic questions and other questions was reported and circulated at a further meeting at which members of each national group were present. In addition, researchers commented on issues of face validity for the questions. As a result of these processes, 5 diagnostic questions were written to be administered in 5 of the 6 countries in the Autumn of 1997. These questions will be termed 'the probes'. In some cases, findings from pilot studies were used to generate closed-response items, broadly following the method of Aikenhead et al. (1987). This process is illustrated for the Surprising Results probe in Section 4. In the following section, details of the probes are presented, together with a critique of the methodology used.

4 The five probes: diagnostic questions designed to investigate the images of science that might affect students' learning during labwork

The relationship between the 5 probes and the 9 research questions is shown in table 2:

TABLE 2: THE RELATIONSHIP BETWEEN THE PROBES AND THE RESEARCH QUESTIONS

	Differences in values from measurement	*Theory and data in scientists' work*	*The nature of scientific results*	*Surprising results*	*Interpreting data*
RQ1	•			•	
RQ2	•				
RQ3					•
RQ4*					
RQ5		•	•	•	•
RQ6*					
RQ7*					
RQ8*					
RQ9				•	

As can be seen, some of the research questions (marked *) were not used in the survey. RQ4 related to students' understanding of the logic of proof and falsification. Previous research suggests that although this may be an issue for younger science learners, it is unlikely to present major problems to students at upper secondary school or beyond (Kuhn et al., 1988; Driver et al., 1996), and RQ4 was not therefore investigated. RQ6 relates to whether activities in the teaching laboratory are seen as involving 'finding out'. Again, previous research suggests that this is unlikely to be problematic at upper secondary school and beyond (Driver et al, 1996), and so the research question was not investigated. In the case of RQ7 and 8, we were unable to develop probes that generated data of sufficient quality to address the questions directly. However, it was possible to gain insights into RQ7 using the Probes in their existing form.

Some of the research questions are only partly addressed by the existing probes.

In the following sections, the design and methodology of each of the 5 probes is presented and discussed. The full text of the probes (in English) can be found in appendix 1. Translated versions can be found in appendix 2. The way in which pilot data were used in the generation of closed responses is illustrated for the probe *Surprising Results*.

Differences in values from measurement

The broad focus of this question was upon students' views about the existence and sources of uncertainties in measured data, and how any uncertainties thought to exist ought to be overcome in selecting a value (RQ1 and RQ2).

Many professional scientists have to make estimates of values from sets of measurements in their work. However, examples of this practice as used in professional science were deemed too complex to be presented to, and understood by, the students in our sample, who are at different stages of their science education and in some cases come from different disciplinary backgrounds. In order to investigate students' views about the errors in measured data, it was therefore decided to present a caricature of the process, involving simple measurements of mass. The question thus involved eliciting students' responses to a laboratory situation. Pilot work using open-response questions suggests that students considered the vignette sufficiently realistic to merit responding to the question.

The probe was set in the context of nutritionists trying to measure the mass of $100cm^3$ of nut oil and soya oil by making several measurements of aliquots taken from a larger sample. [Of course, there are better methods of determining the mass of $100cm^3$ of oil than this and professional nutritionists would no doubt use them.] In the first instance, students were presented with two sets of 9 measurements of mass, arranged in ascending order, and with a mean value. The two sets of measurements were presented as having been made by two different groups from the same original sample of oil. The spread of measurements from one group was wider than from the other group, though both data sets had the same average. The average value corresponded to a measured value on one list. One measured value was repeated in one list, and this was different from the middle value. All values and averages were given to 3 significant figures.

Students are asked to write down what the nutritionists should conclude to be the mass of $100cm^3$ of oil. Pilot work suggested that some students select the mode (presumably on the grounds that the 'true' value must be one of the measured values; Séré et al., 1993). Others select the average, presumably indicating some understanding of the use of a *set* of measurements to arrive at an estimated value. Others select the average with some notion of confidence limits, or omit extreme data points, indicating that the value is seen as an estimate. It was also envisaged that some students might make open comments about the methodological approach used by the nutritionists.

Students were then asked to choose one of six possible statements relating to the confidence that could be attributed to findings from the two sets of measurements. The six statements were derived from previous research and piloting. The responses refer to the existence of repeat measurements, and the relative spread of the data sets. It was intended that this part of the question would provide further insights into how students arrived at a value for the mass of $100cm^3$ of oil.

The next part of the question was similar in format, only this time the students were presented with two sets of 9 measurements of the mass of $100cm^3$ of a different oil. The design of these data sets was rather different, in that the average values were different and neither average corresponded to a measured value. The difference in averages was small compared to the spread of the data. One value was measured by both groups. The spread of one data set was wider than that of the other.

Students were asked similar questions in order to see how a value was arrived at from the available data.

In the last part of the question, students were presented with 2 further data sets. The first consisted of 9 measurements of the mass of $100cm^3$ of one oil, the second consisted of 9 similar measurements for another oil. The mean values were different. Students were asked whether the mass of $100cm^3$ of the two oils was different, and were asked to select from 5 closed-response answers which had been derived from previous research and piloting. These closed-response statements referred to the spread of the data sets, the existence of repeated measurements and the size of the difference between the averages compared to the spread of the data set.

The final form of this question was written by Robin Millar and Marie-Geneviève Séré, following a period of discussion within the project. The question draws heavily on previously published work (Lubben and Millar, 1996; Séré et al., 1993).

The question was not piloted in its final form, before being used for data collection. It is recognised that some students may view the spread of the data sets with suspicion, the implication being that the errors are excessively large. Indeed, a small number of students in the pilot study did comment on the spread of data, suggesting that this indicated incompetence on the part of those carrying out the measurements. Our primary interest, however, was in how students select values from sets of measurements and the closed-response question does not therefore provide opportunities for students to comment on the methodology used. These factors may result in some students being unable or unwilling to respond to the question, and this possibility was borne in mind during analysis.

Theory and data in scientists' work

The broad focus of this question is upon how students conceptualise relationships between theory and data, and in particular whether students entertain the possibility of more than one interpretation of a given data set (RQ5).

The question was written to try to gain insights into the general epistemological commitments of students, and it was therefore decided not to suggest a context. At the analysis stage, it is hoped that responses of individual students on this question will be compared with their responses to similar contextualised questions.

Students were presented with 7 pairs of statements about relationships between scientific theories, empirical data, and the design of investigations. The statements were written to express opposing viewpoints, though they cannot be viewed as logical opposites. Students have to rate their position on a 5 point Likert scale. Notes about the use of the Likert scale were provided, the middle point indicating no strong opinions OR that some merit is seen in both statements. Some pairs of statements attempted to probe similar issues using different wording. For example, pairs 1 and 4 both relate to experimental design, pairs 2 and 6 relate to data analysis.

It is hypothesised that students' responses on the Likert scale will provide insights into their views about the relationships between theory and data in investigations. Students who view theory as emerging from data are likely to respond differently from students who see theory and data as interrelated.

The final form of the question was written by Milena Bandiera and Hans Niedderer, with some contributions from John Leach. It is based on pilot questions used in Germany and Italy.

It is hoped that students' responses to this question will give insights into their general epistemological orientation, and responses to this question will be compared with contextualised questions with a similar focus. A number of criticisms of the methodological approach can be made, however. As the question is decontextualised, a different pattern of ticks may well emerge if different examples of scientific theories and investigations are considered. To exemplify this, consider the pair of statements 'One data set always leads to one conclusion/Different conclusions can legitimately result from the same data'. For many data sets, it is hard to imagine more than one conclusion that might legitimately be drawn. For example, it is hard to make a legitimate case for more than one answer for the position of a moving object, using data relating to its initial position and velocity. However, for other data sets there may well be a range of different legitimate conclusions. For example, various accounts could be given to explain data about the intracellular concentration of a metabolite. An additional problem relates to how the paired statements are interpreted. For example, the statement 'Scientists interpret data without being influenced by their theoretical assumptions' might be interpreted as 'Scientists try to minimise bias in their work', rather than 'Scientists do not draw upon *a priori* theory when analysing data: the data speak for themselves'. Indeed, pilot work suggested that both these interpretations were indeed used.

The nature of scientific results

This probe also addresses students' understandings of the relationships between theory and data (RQ5). However, questions are answered by students in the context of a stated knowledge claim.
Students are presented with 3 statements of reliable public knowledge, relating to physics, chemistry and biology. These were termed *SCIENTIFIC STATEMENTS*. One *SCIENTIFIC STATEMENT* has to be selected, and then students are asked to record their opinions about how the *SCIENTIFIC STATEMENT* was first put forward, and why it is believed, on Likert scales. Closed responses explaining how the *SCIENTIFIC STATEMENT* was put forward mention the role of data, imagination, theoretical and mathematical thinking, accumulation of knowledge, and logical arguments and deductions. Closed responses about why the *SCIENTIFIC STATEMENT* is believed refer to similar issues, together with the standing of the researcher who proposed the *SCIENTIFIC STATEMENT*. The closed-responses were written drawing upon issues in the history, philosophy and sociology of science, and previous research on students' epistemological understanding. Similar diagnostic questions were piloted in a closed response format, in a number of countries.

The final formulation of this question was produced by Milena Bandiera and Hans Niedderer, with some contributions from John Leach. It drew upon questions piloted in Italy and Germany.

The wording of the closed-responses used in this question was not piloted prior to data collection, and it is therefore possible that unintended meanings were communicated to students. For example, some students may agree strongly that their statement was first put forward as a result of data which were collected through accurate and repeated experiments, on the grounds that accurate and repeated experiments were indeed carried out. Others may disagree with the statement on the grounds that the statement was put forward before empirical work was carried out. Others may have no strong opinion, assuming that the processes of empirical work and proposing theoretical statements are related. Some may think that although data were collected, this is not why the statement was put forward. Students may not make distinctions between ‘theoretical and mathematical thinking’ and ‘logical arguments and deductions’. Bearing these issues in mind, findings from analysis have to be treated with some caution.

Surprising results

This question has a number of interrelated foci, relating to students’ ideas about relationships between theory and data, about possible sources of uncertainty, and about the influence of the status of knowledge claims upon the treatment of data (RQs1, 5 and 9).

The primary focus of the question, however, was upon how students think that the status of knowledge claims under investigation within the scientific community might influence the actions that are appropriate in investigations (RQ9). In order to probe this, it was decided to present students with two examples of investigations where different groups reached different conclusions. In one case, the knowledge claim being investigated was agreed as reliable amongst scientists (starch is produced in plant leaves in the light, but not in the dark). In the other case, there was no agreed position within the scientific community (there is a link between leukaemia clusters and the siting of chemical plants). This probe therefore elicited students’ responses to laboratory (and other) investigative situations.

Firstly, students were presented with a description of some labwork in a school classroom. Students were testing leaves for the presence of starch, from plants kept in the light and in the dark for several days. The students’ results were inconclusive: some got results agreeing with the textbooks, others did not. Students were asked what they would do in this situation, as the teacher of the class. Six closed-response statements were written from pilot data using open-response questions (as described below). These statements referred to collecting more data or data of better quality, drawing upon theory to explain how the unexpected results might have come about, referring to over-simplifications of theory in textbooks, and abandoning textbook theory on the grounds of the students’ results. Each statement had to be rated ‘Agree’, ‘Disagree’ or ‘Not sure’ by the students. Finally, the students had to state which one course of action they thought would be best.

In the pilot studies, it seemed from the open responses of students that some believed that empirical data always leads to clear solutions. The question used in the pilot study presented questions in an open format, students being asked to answer in their own words 'What should the teacher do in this situation?/What should the government ministry do now?' and 'Explain why you think this is the best thing to do'. Finally, they were asked to answer in their own words 'If your answer in Part 2 was different from your answer in Part 1, please explain the reason for this'.

The open responses given by students in GB (n = 36) and France (n = 105) were read through independently, and broad categories of students' responses were identified. In addition, insights from the German pilot and some detailed interviews with small samples of students carried out in Denmark were drawn upon. For part 1, where students were asked what the teacher should do now, responses referred to three main areas. There was a surprising degree of consistency in the categories of responses identified in each country. The first of these broad areas stated that the teacher's actions should involve collecting more data from which an answer would become clear (GB: 19 students; France: 30 students). Responses referred to eliminating errors and having more data to draw upon, with no mention of the underlying theory being investigated. The second main area of responses referred to links between data and theory in explaining what the teacher should do (GB: 16 students; France: 57 students). Students suggested that the teacher should explain what happened in terms of theoretical explanations of photosynthesis, or in terms of the experimental methodology used, so that students would understand the accepted theory but also know why the application of theory to real contexts can be problematic. The final main area of responses referred to students' preferred teaching styles rather than relationships between data and theory (GB: 9 students; France, 47 students). A very small number of responses referred to other issues.

These findings were drawn upon in generating closed response items. The main problem faced was how to write a manageable number of statements which reflected *both* what students thought the teachers should do *and* the students' reasons for this. In the end, a number of two-part statements were written, which state *both* a course of action for the teachers *and* a rationale for that course of action. For example, the following statement was written to capture the views of some of the students who stated that a result should become clear in an unproblematic way: 'Repeat the experiment to check the results: if the students work carefully enough, the expected result will become clear.' 6 such statements were written, and in addition space was provided for students to write other responses in their own words.

The closed response items used in this question were not piloted before the main data collection, and it is therefore possible that unintended meanings will be communicated to students. In addition, as each statement has two parts there is a possibility that students will agree with one part of the statement but not the other.

A similar process was undertaken for Part 2, though this question was not piloted outside GB. Students were asked to explain what the government ministry should do next. Responses were quite similar to those given in Part 1, referring to the methods used by the scientists (GB: 6 students), the data available (GB: 15 students), and the underlying mechanisms that might explain possible links (GB: 7 students). A very small number of responses referred to other features.

Seven two-part closed response statements were written to capture the above viewpoints, and in addition students were given the opportunity to present other possibilities in their own words in the final wording of the probe.

Few students explained why responses to Parts 1 and 2 were different, most giving a legitimate response that the two situations are completely different. No attempt was therefore made to produce closed response statements for this part of the question, which was dropped in the final version of the probe.

The final formulation of this question was worked on by Dorte Hammelev, John Leach, Jenny Lewis and Robin Millar, and Albert Christian Paulsen. The original version of the question was written by Robin Millar.

Interpreting data

The focus of this probe is upon whether students see processes of suggesting relationships between data points as routine heuristics, or whether such processes are seen as more similar to hypothesising (RQ3). In order to investigate this research question, it was decided to choose a context in which data points were being measured in order to investigate a theoretical model linking physical quantities, and the process of linking data points was therefore a process of hypothesising.

In the pilot study, the conductivity of a new material was selected as the context, and relationships about voltage and current were presented to students. However, although this context was presented in a non-technical way many upper secondary students who were not studying physics nonetheless interpreted the question as relating to physics and were therefore unwilling to answer it. In addition, the responses of some physics students were clearly influenced by what was thought to be a correct answer to the question without any reference to the presented data. Attempts to develop comprehensible questions in the context of enzyme kinetics and models of climate change were unsuccessful. The context finally developed is based on models of superconductivity, drawing upon contemporary work in solid state physics.

The question is designed to find out whether students think that it is possible to tell which is the better of two competing models on the basis of empirical data, or alternatively whether students think that consideration of the underlying models has a place. In the first instance students were presented with a diagram which presented the resistance of a material at various temperatures. All data points were presented with error bars. It was stated that although this data set had been collected by one research group, its form was broadly agreed by a number of research groups at a conference. However, the groups at the conference did not agree about the relationship between resistance and temperature, because two different theoretical models were being drawn upon. These different theoretical models resulted in two different lines being used to link the data points. Both lines lay within the error bars for the measurements. Students were then presented with two diagrams showing the different proposed links between temperature and resistance. [The theoretical models underpinning the proposed relationships between temperature and resistance are highly technical, and were not presented.] The question therefore presents a critical laboratory incident to students.

Students were asked to select one of six statements about the two lines. These statements were derived from responses to the pilot question (although the context of the pilot question was not superconductivity). The statements referred to choosing one or other line on the grounds of empirical consistency, collecting more data so that one or other line could be chosen on the grounds of empirical consistency, rejecting both lines, looking at the underlying models from which the lines were generated, and allowing each group to choose which line to use as there is no way of judging between them. In addition, students were allowed to suggest another position in their own words, if none of the presented positions matched their own viewpoint.

The next part of the question asked students to rate 8 statements about what the scientists should do next as appropriate, inappropriate, or not sure. Again, the statements were derived from pilot work using a different context, and referred to various strategies involving drawing conclusions based on data alone (or bigger data sets with fewer errors), considering the underlying models, looking for other underlying models, and leaving the choice of models up to individual scientists as there is no way of finding out which interpretation is the correct one. In addition, students were given the option of suggesting a course of action in their own words. Finally, students were asked to choose one course of action as the best one for the scientists to follow.

The final part of the question related to discussions at the conference about the conductivity of a different material. Again, the general form of a data set was agreed, but different groups of scientists had different views as to how the data points should be linked. Students were presented with four closed statements as to what should be done next, derived from pilot work. These referred to joining each data point with a line as each measurement is known with confidence, considering underlying models, using a computer to generate a line of best fit, and allowing each group to make their own decision as there is no way of knowing which is the best line. Students were also given the opportunity to make an open response of their own.

Throughout this question, it might be expected that students who see the joining of data points as a routine heuristic would suggest making a decision on the grounds of the data points alone, whereas those that see the process as one of hypothesising would suggest looking at the underlying models. In addition, choices are presented to allow students to express the view that there is no way of judging between the lines.

The rationale behind the choices offered to students is presented in table 3:

TABLE 3: RATIONALE BEHIND CLOSED RESPONSE CHOICES OFFERED

STUDENTS' POSITION	*FIXED-RESPONSE ITEM WHERE THIS VIEW COULD BE EXPRESSED*		
	Part 1	Part 2	Part 3

It is possible to establish which group is correct on the basis of data alone. More (or better) data will make it clear who is correct.	A, B, D, F	A, B, C, D, F,	DidaScO TESME
To establish the correct explanation of the data we must consider the models which lie behind the lines drawn through the data.	C	E	BREM
Anything goes. There is no way of knowing who is correct. All interpretations are equally valid.	E	H	ROMA
Other	G	I, G	

The final formulation of this question was produced by John Leach, Jim Ryder and Vasilis Tselfes, with some contributions from Robin Millar. It draws upon a number of draft questions written in several countries, notably Greece and Great Britain.

The context used in this probe and the wording of closed response items were not piloted prior to data collection. It is therefore possible that these communicated poorly to students. It is also recognised that this probe presents a vignette of discussions between scientists at an academic conference, in order to highlight specifically the grounds on which choices between different interpretations of a given data set might be made. At the analysis stage, it was therefore important to try to make judgements about the face validity of the context to students answering.

General methodological issues

In the previous sections, a number of methodological concerns have been raised as to whether it will be possible to elicit useful information about students' images of science using the probes. In this section, some more general points about the methodology used in the study are made, as well as the types of claims about students' images of science and labwork that might be supported by the study.

In designing written, closed-response questions to investigate students' images of science, it has been necessary to make a number of difficult choices. One of these relates to simplifying the contexts used in the probes so that they can be understood by upper secondary students as well as students who have never encountered the context before, while maintaining the face validity of the questions in the eyes of more senior university students with more first-hand experience of the context. The probes present vignettes of small parts of the work of some groups of professional scientists, and it is unlikely that any professional scientist would say that her/his work is accurately reflected by these vignettes. However, it is hoped that some professional scientists would recognise some of their practices in the probes, such as selecting a value from a set of measurements, judging the quality of a set of measurements, debating the relative merits of models in the context of data, and so on. In presenting such vignettes to students, it was intended to gain some insights into their images of science. The main issue of face validity for the probes relates to the how they are perceived by the students who respond to the questions, and the extent to which students' responses are understood by the researchers, rather than the extent to which professional scientists view the contexts as authentic. The construct validity of the probes relates to the extent to which the practices presented relate to activities undertaken by students during labwork, or activities undertaken by professional scientists in their work. It is, of course, possible that the vignettes of the work of scientists used present scientific practices in a way that does not allow inferences to be made about students' images of science. This issue is discussed in the light of the data collected in the study.

There is a good deal of evidence from the pilot study that the contexts presented to students do indeed yield insights into their images of science, even though the contexts present vignettes of the work of professional scientists. For example, in responding to the pilot version of *Interpreting data*, some students referred explicitly to an interplay between theoretical models and data in empirical enquiries. This was taken as evidence that the probe was successful in allowing such students to express an epistemological position. Some students' responses on the same question referred only to data. This might indicate a view that theory emerges from data. However, it is also possible that such responses indicate a lack of understanding of the context as presented on the part of such students and this was borne in mind at the stage of interpretation of results. However, similar epistemological positions were noted across responses to several pilot questions, indicating that the effect is not restricted to students' understanding of one context.

The use of closed responses was deemed necessary as the probes are to be administered to a multi-language sample. However, this has serious implications for the level of detail at which students' understandings can be investigated. The wording of the closed responses constrains the types of things that students can say. Students who are used to the norms of schooling, where students answer questions posed by teachers no matter how implausible the questions seem, are likely to tick a box whether they agree with the statement or not. During the GB pilot, the same survey was presented to 5 students on 2 occasions, separated by several weeks, to find out the extent to which similar viewpoints were expressed. This limited evidence suggested that students did indeed respond to the open questions with similar viewpoints, though the closed-response items generated have not been piloted and retest reliability has not been measured.

At the stage of analysis, researchers have little to go on in interpreting what students understood by the contexts of the probes, and the wording of the closed response items. A small-scale interview study would have given useful insights into what is meant by ticks in particular boxes, though it was not possible to conduct such a study within the lifetime of the project *Labwork in Science Education*. However, it is possible to use findings from some of the Case Studies to evaluate the extent to which students draw upon similar images of science in responding to the 5 probes, and in carrying out labwork. This issue is discussed later in the paper.

For a closed response instrument, the final reporting categories are constrained by the closed response items presented. The validity of the research instrument therefore depends upon the extent to which the closed response items are understood by members of the sample in the way that was intended by the researchers, and that students are able to express their viewpoints in terms of the items available. The generation of closed response items is therefore of critical importance in the methodology of the survey. The closed response items for the 5 probes were generated from available pilot data, as well as researchers' knowledge of previously published work. However, the evidence used to generate some of the closed response statements was limited, and it is possible that some students' preferred responses are not reflected in the statements. In addition, the statements generated for *Theory and data in scientists' work* and *The nature of scientific results* were generated more from the researchers' insights into the history, philosophy and sociology of science, than from pilot data from students at the upper secondary and university levels. It is therefore possible that students will not be able to express their viewpoints in terms of the closed response items presented.

We have previously expressed some reservations about the use of decontextualised questions in eliciting insights into the images of science that students might use during labwork. The work of Samarapungavan (1992), for example, shows how professional scientists espouse different views about the nature of science in different contexts, more elaborated positions being articulated about specified, well-known contexts. This can also be seen in pilot responses in this study. As stated earlier, the pilot survey was administered to a small number of teachers. In GB, the upper secondary teachers used were all very experienced in the use of labwork and their responses in general were thoughtful and sophisticated. Different positions about issues such as relationships between theory and data were stated in different contexts. For example, one pilot question asked teachers to state their position about the following statements:

- Good scientists ought to have theoretical assumptions which influence their analysis of data;

- Good scientists ought to be neutral and objective: it is not acceptable for scientists' theoretical assumptions to influence their analysis of data.

One biology teacher stated disagreement with the first statement and agreement with the second, adding the comment 'Why bother collecting data if you are God?' in explanation. This could be taken as indicating the view that theory is derived from data. However, the teachers' responses to contextualised questions indicated a different story. In explaining what should be done next on Part 1 of *Surprising results*, the teacher suggested that a lesson should be given on 'the imprecision of science' (which was taken to mean the problematic relationship between the behaviour of natural phenomena and theoretical predictions about how phenomena will behave). Similar points were made on Part 2 of the question.

For all the reasons described above, claims about students' images of science arising from the study are presented cautiously.

5 Details of sample

Data collection took place in five countries during the Autumn of 1997. Table 4 shows the sample size in the participating countries:

TABLE 4: SAMPLE FOR MAIN DATA COLLECTION

COUNTRY	*UPPER SECONDARY STUDENTS* TOTAL *Phys. Chem. Biol.*			*UNIVERSITY STUDENTS* TOTAL *Phys. Chem. Biol.*			*TEACHERS* [Upper secondary *University*]
Denmark [1]	219			0			39
	88	*47*	*84*	0	0	0	*0*
France [2]	103			70 [4]			0
	103	103	103				*0*
Germany [3]	55			84			0
	15	24	16	32	29	23	*0*
Great Britain [3]	82			85			2
	33	38	31	26	25	34	*4*
Greece [3]	78			60			0
				22	15	23	*0*

Total upper secondary students for *Differences in Values from Measurement* and *Surprising Results*: 422
Total upper secondary students for *Theory and Data in Scientists Work, The Nature of Scientific Results* and *Interpreting Data*: 432

[1] In the Danish sample, each respondent only answered a subset of the questions.
[2] All science stream students study each science subject in France.
[3] Many students in these samples studied more than one science subject, but not necessarily all three.
[4] French university students study some physics, chemistry and biology during the first two years, even though their courses ultimately lead to different specialisms. The sample has been constructed to include students working towards different specialisms.

In addition, an extensive study of teachers' images of science was conducted in Italy (see Bandiera et al., *Working Paper 4*).

It was intended that the sample in each country, at each level, would include between 50 and 100 students, at least 30 of whom would be studying biology, chemistry and physics. In practice, students in each of the participating countries study more than one of these subjects at the upper secondary level, and in some cases at the university level. At the upper secondary level, it is not therefore legitimate to report responses from physics, chemistry and biology students separately. By the university level, however, students in each country have committed themselves to some degree of specialisation into the disciplines of physics, chemistry and biology. Students' responses are therefore reported by discipline in all countries except France, where specialisation in teaching does not occur to any great extent until later in degree courses.

At the upper secondary level, the sample in each country was drawn from a minimum of three schools and was constructed to be as representative of students in academic strands of education as possible[2] . The sample for university students was constructed to be as representative of full cohorts as possible. The sample cannot be viewed as representative at the national level, however, because the sample was not designed to

[2] All survey samples in the project focus upon academic streams. In some countries the distinctions between 'academic' and 'vocational' streams are very rigid (e.g. France) whereas in other countries the system is more flexible (e.g. GB).

be representative of the diversity of students and institutions at national level. The study should be viewed as a small scale exploratory study of European students' images of science in labwork learning, rather than a comparative study. In addition, the survey was administered in the national language of each country. It cannot be assumed that the same meanings were broadly inferred from each question, in the different language versions. Findings from the study are not therefore reported country by country, though attention is drawn to particularly large differences between samples in different countries.

The probes used were designed around hypotheses relating students' images of science to their learning during labwork, and as such they are not necessarily appropriate tools for investigating the images of science drawn upon by teachers during labwork teaching. A small sample of responses were nonetheless collected from teachers. The purpose of this data set is to consider the communicability of the questions, the assumption being that the reading and understanding skills of teachers should allow them to interpret the questions as written. The aim is not to characterise teachers' images of science. Rather, this information will be used to consider possible differences between teachers' and students' understandings that might be relevant to communication between teachers and students during labwork.

It has been assumed that if the Probes communicate clearly, then teachers in the sample ought to respond by drawing upon reasonably sophisticated images of science. In section 6, qualitative descriptions of teachers' responses are therefore presented and conclusions are drawn about the communicability of the Probes.

6 Analysis and findings

In the first instance, data analysis consisted of frequency counts of the closed response choices selected by students, plus simple coding of open-response items.

Earlier in this paper some of the methodological issues surrounding the contextualisation of the probes were discussed. In the end, some of the probes used address broadly similar research questions, but with different degrees of contextualisation in the design of the probe. For example, the probes *Theory and Data in Scientists' Work*, *The Nature of Scientific Results*, *Surprising Results* and *Interpreting Data* all relate to the fundamental issue of perceived relationships between theoretical knowledge claims and empirical data. In *Theory and Data in Scientists' Work*, questions are presented with no information about the context in which experiments are performed or data are collected, whereas in each of the other probes a context is suggested and, in some cases, elaborated in some detail. Data were therefore examined to assess the extent to which individual students' responses are dependent upon the contextualisation of the probe, in order to comment upon the extent to which students use different images of science in specific contexts, as opposed to drawing upon general epistemological orientations across a broad range of scientific contexts.

In some cases, students marked more than one response category even though they were asked only to mark one. All such responses were coded, giving the possibility of frequency counts greater than 100%. However, the population size for each part of each probe was not amended due to null responses, and most frequencies do not therefore total 100%.

6.1 Differences in Values from Measurement

6.1.1 SAMPLE FOR THIS PROBE, AND REPORTING CATEGORIES

The sample responding to this probe is given in Table 5:

TABLE 5: SAMPLE FOR THE *DIFFERENCES IN VALUES FROM MEASUREMENT* PROBE

Country	*Upper secondary*	*University physics*	*University chemistry*	*University biology*
Denmark	104	0	0	0
France	103	*	*	*
Germany	55	32	29	23
Great Britain	82	26	25	34
Greece	78	22	15	23
TOTAL	422	80	69	80

*Sample not divided by discipline. 70 students in total.

Data are reported as frequency counts, expressed as a percentage of the total sample. Data from upper secondary students have not been divided into physicists, chemists and biologists. Data from University students is reported by discipline in all countries except France. Frequencies for the French university sample are reported separately.

The three parts of the Probe are reported separately.

6.1.2 TABULATION OF DATA

Two data sets, one oil, same average: What should group A state as their result for the mass of 100 cm^3 of nut oil?

Two data sets of measurements of the mass of 100cm^3 olive oil were presented, each containing different measured values, but yielding the same mean value. Students were asked to state the mass of 100cm^3 olive oil, and to justify their reasoning. They were then presented with six statements about the relative confidence that could be ascribed to the two data sets. They were asked to select one statement that they most closely agreed with.

Students' responses were coded into the following categories:

87.1g
Other

[It is interesting to note that hardly any students stated the mass to two significant figures.]

The justifications offered by students were then coded into broad categories. The English responses were coded in English. The French responses were coded in French. Other responses were translated into English and then coded. The German responses were not coded. The categories used for coding were as follows:

A It is the average/the best estimate
B The need to state confidence limits mentioned
C The possibility of deleting highest and/or lowest measurements mentioned
D Suggestion that 87.1g was the most common measurement
E Suggestion that the value 87.1g was closer to the 'real' value.
O Other

A code was allocated for each suggestion. So, if a student stated '87.1 ± s.d. is the average plus or minus the standard deviation', a code of A and B would be allocated. In practice, not all of the justifications in the multilingual sample were coded.

Students were then asked to tick one statement A-F that they most closely agreed with. The statements compared the two data sets. In practice, some marked more than one statement. All students' statements were coded, and the percentages therefore total more than 100.

Frequency counts are summarised in Table 6:

TABLE 6: THE MASS OF 100cm^3 NUT OIL, WITH JUSTIFICATIONS - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary	University physics	University chemistry	University biology	French University	**TOTAL: University**
Average = 87.1g	85	83	88	90	90	86
Average = other	12	15	14	6	7	10
Justification A: It is the average	53	*	*	*	39	*
Justification B: Need for confidence limits mentioned	14	*	*	*	15	*
Justification C: Deleting extremes mentioned	5	*	*	*	0	*
Justification D: 87.1 most common measurement	3	*	*	*	0	*
Justification E: 87.1 closer to 'real' value	3	*	*	*	7	*
Other justification	8	*	*	*	4	*
Agree with A: Range greater in group A	2	1	1	1	0	1
Agree with B: Group A get repeat values	6	4	7	5	0	4
Agree with C: One of Group A's measurements was 87.1	5	0	6	1	1	2
Agree with D: Range in Group B's	50	56	72	50	67	61

measurements less						
Agree with E: Equal confidence in each group's measurements, as the average is the same	36	30	16	46	13	27
Agree with F: Can't decide which group's result can be used with greater confidence	20	15	19	23	20	19

* Not analysed

By far the majority of all students selected 87.1g as the mass of 100 cm^3 olive oil. Students suggested this for a variety of reasons. The most popular statement selected was D, apparently indicating some understanding amongst students that the spread of measurements in a group gives some indication about the confidence that can be attributed to a calculated average. More university students selected this than upper secondary students (61% compared to 50%). However, the next most popular choice was statement E, possibly suggesting that students judged the confidence that can be ascribed to the estimated mass in terms of the actual value selected as opposed to the measured data on which the estimate was based. A small but significant number of students who chose either statement B or C suggested some notion that the 'real' mass of oil must correspond to a value that was actually measured (11% at upper secondary; 6% at university).

Teachers responding to this question overwhelmingly selected the presented statements D and F, indicating that the question communicated to them sufficiently to indicate that the issue of interest is the implications of the spread of data for confidence in any estimate made. A small number of upper secondary teachers selected statement E, however.

Two data sets, same oil, different averages: What would you state as the mass of 100 cm^3 of soya oil?

The students were presented with two sets of data for soya oil, each set having a different average and spread. They were asked to state the mass of 100cm3 soya oil, and to explain how they arrived at their result. Finally, they were presented with five statements about the relative confidence that could be ascribed to the two data sets. They were asked to select one statement that they most closely agreed with.

Students' responses were coded in a similar way as on the last part of the question. In this case, however, students suggested a range of possible values for the mass of 100 cm^3 of soya oil. These were:

83.75g (i.e. the average of data from Groups A and B, when totalled)
83.7g
83.4g (i.e. Group A's average)
83.8g (i.e. a value measured by both groups A and B)
Other

Again, students did not tend to state a mass to two significant figures. The justifications offered by students were treated in a similar way to the last part of the question, and were coded into the following categories:

A Average of all data when combined
B A value measured by both groups A and B, or near to actual measured value
C Group A's data has less spread, so this is used as the data set
D The median value
O Other

Frequency counts are summarised in Table 7:

TABLE 7: THE MASS OF 100cm^3 SOYA OIL, WITH JUSTIFICATIONS - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary	University physics	University chemistry	University biology	French University	**TOTAL: university**
Average = 83.75g	38	19	38	54	37	36
Average = 83.7g	3	10	6	0	9	6
Average = 83.4g	23	25	30	23	31	27
Average = 83.8g	15	21	13	6	12	13
Average = Other	12	13	12	8	7	10
Justification A: Average of full data set	49	30	42	53	53	44
Justification B: Measured by both groups/near to actual measured value	8	8	1	1	1	3
Justification C: Group A's data used due to smaller spread	19	13	22	19	17	17
Justification D: Median	1	0	0	1	1	1
Justification E: Other	10	5	1	5	3	1
Agree with A: Range less in group A	42	50	57	43	57	50
Agree with B: Range greater in Group B	3	1	3	1	1	2
Agree with C: Equal confidence	9	6	4	8	1	5

in each group, as both groups measured 83.8g						
Agree with D: Equally confident in each group, as averages close	20	16	12	30	16	18
Agree with E: Can't decide which group's result can be used with greater confidence	35	33	33	29	26	29

A similar picture emerges on this part of the Probe as on part 1. Although most students selected statement A, indicating an understanding that the spread of measured data influences the confidence that can be attributed to a calculated estimate, a small but significant number suggested that the *value* of the calculated estimate is the critical factor (i.e. statement D; 20% of upper secondary, 18% of university), or that the 'real' value should correspond to an individual datum (i.e. statement C; 9% of upper secondary, 5% of university).

Teachers overwhelmingly selected statements A and E, indicating that the question was interpreted in the manner intended. A small number of upper secondary teachers did, however, select statement D.

<u>*Two data sets, two oils, similar but different averages: Comparison of masses of sunflour oil and olive oil*</u>

In this part of the question, students were presented with two data sets relating to two different oils. The averages were different, but close compared to the spread of the data sets. Five statements were presented relating to whether the oils did indeed have different masses. The students were asked to mark the one statement that they most closely agreed with. Frequency counts are summarised in Table 8:

TABLE 8: IS THE MASS OF 100cm^3 SOYA OIL DIFFERENT FROM THE MASS OF 100 cm^3 OLIVE OIL? - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary	University physics	University chemistry	University biology	French University	**TOTAL University**
Agree with A: the mass of olive oil is greater, as the average is greater	29	13	33	28	16	22
Agree with B: There is no difference, because the range of measurements is much greater than the difference between averages	8	10	3	9	6	7
Agree with C: No difference, because 94.1 measured by both groups	4	1	0	5	1	2
Agree with D: We cannot be sure, because the range of measurements is much greater than the difference between averages	47	71	68	26	53	62
Agree with E: We cannot be sure, because 94.1 measured by both groups	17	7	6	10	16	10

Data from this part of the question is particularly interesting. Statements B and D both refer explicitly to the spread of data; 55% of upper secondary students and 69% of university students selected one of these two statements. However, 29% of upper secondary students, and 22% of university students selected statement A, implying a focus on the value of the calculated average rather than upon the data from which the value was calculated. There are also interesting subject differences amongst the university students selecting response A (13% of physicists, 33% of chemists, 28% of biologists) and response D (71% of physicists, 68% of chemists and 26% of biologists). 17% of upper secondary students and 10% of university students selected statement E, again indicating a view that the 'real' value of the measurement ought to correspond to a measured datum.

Teachers' responses were more spread on this part of the Probe. Although no teachers selected statement C, small numbers selected statements A and E. The vast majority, however, selected statements B and D indicating that the question communicated broadly as intended.

6.1.3 DISCUSSION OF RESULTS

General comments

Most students in the sample appeared able to follow some sort of algorithm to select a value from a set of measured data. For some students, this algorithm appeared to be underpinned by explicit understanding of the implications of spread of data for the confidence that can be ascribed to an estimate (e.g. students choosing statement D on part 1, A in part 2 and B or D in part 3). Around 40 - 60% of students in the sample were in this category, with about 10% more students at the university level responding in this way than the upper secondary level.

However, for many students, the use of algorithms in selecting values did not appear to be underpinned by an explicit understanding of the implications of spread for confidence (e.g. students choosing statement E in part 1, statement D in part 2 or statement A in part 3). About 20 - 35% of students in the sample were in this category, with little evidence of difference between the university and school levels.

Between about 5 and 15% of students in the sample appeared to hold a view that 'real' values correspond to measured data, selecting statement A or C in part 1, C in part 2 and E in part 3. Although the numbers are small, marginally fewer university students appeared to reason in this way.

The teachers' responses to the question suggested that, broadly speaking, the question communicated as intended.

Subject-related differences at the University level

Examination of Tables 6, 7 and 8 suggests that university biology students have less appreciation of the implications of spread of data for the confidence that can be attributed to estimates of values, and that biology students are marginally more likely to hold the view that 'real' values should correspond to measured data. This is surprising, in that much data handled in biology takes the form of frequency counts (e.g. population data in ecology, assay results in biochemistry), and this data is often subject to some form of statistical analysis in order to draw conclusions. However, in some countries in the sample relatively little attention is given to the spread of data during biology teaching in the first year of university (e.g. Greece). A possible curriculum goal for labwork in all three sciences, and particularly biology, might therefore be to teach students about confidence, accuracy and precision in data handling.

6.2 Theory and Data in Scientists' Work

SAMPLE FOR THIS PROBE, AND REPORTING CATEGORIES

The sample responding to this probe is given in Table 9:

TABLE 9: SAMPLE FOR THE *DIFFERENCES IN VALUES FROM MEASUREMENT* PROBE

Country	*Upper secondary*	*University physics*	*University chemistry*	*University biology*
Denmark	114	0	0	0
France	103	*	*	*
Germany	55	32	29	23
Great Britain	82	26	25	34
Greece	78	22	15	23
TOTAL	432	80	69	80

*Sample not divided by discipline. 70 students in total.

Data are reported as frequency counts, expressed as a percentage of the sample. Data from upper secondary students have not been divided into physicists, chemists and biologists. Data from University students is reported by discipline in all countries except France. Frequencies for the French university sample are reported separately.

TABULATION OF DATA

Students were presented with seven pairs of statements about the relationship between theory and data in scientists' work. The statements were written as far as possible to represent opposite points of view. In each case, they had to mark a scale from 1 to 5, indicating their viewpoint.

1 indicated a strong agreement with statement A, or that it is usually true. 2 indicated agreement with statement A, or that it is generally true. For the purpose of reporting, responses in these two categories have been conflated.

3 indicated no strong opinion either way, or that both statements have some merit.

5 indicated a strong agreement with statement B, or that it is usually true. 4 indicated agreement with statement B, or that it is generally true. For the purpose of reporting, responses in these two categories have been conflated.

Frequencies are presented in Tables 10 and 11:

TABLE 10: FREQUENCY OF UPPER SECONDARY STUDENTS' RESPONSES ON PAIRED STATEMENTS AS % OF SAMPLE

Statement A	1 or 2	3	4 or 5	*Statement B*
The design of an experiment is dependent on theory about the thing that is being investigated.	53	21	24	An experiment is designed to see what happens, and does not depend on theory about the thing that is being investigated.
In analysing a given data set, it is quite reasonable for different scientists to use different theoretical perspectives.	78	9	11	In analysing a given data set, there is only one theoretical perspective which it is reasonable for scientists to use.
Scientists interpret data without being influenced by their theoretical assumptions.	19	17	62	Scientists' theoretical assumptions influence their interpretation of data.
Scientists' ideas and theories influence their planning of data collection in experiments.	59	16	24	Scientists put their ideas and theories to one side when they are planning data collection in experiments.
One data set always leads to one conclusion.	15	13	70	Different conclusions can legitimately result from the same data.
Scientists plan their data analysis based on the ideas and theories that they had when designing the experiment.	67	18	15	Scientists plan their data analysis without reference to the theories that they may have had when designing the experiment.
It is *not always* possible to tell which is the most powerful of two competing theories, no matter how much data are available.	54	23	23	It is *always* possible to tell which is the most powerful of two competing theories if enough data are available.

TABLE 11: FREQUENCY OF UNIVERSITY STUDENTS' RESPONSES ON PAIRED STATEMENTS AS % OF SAMPLE

Statement A	1 or 2 Phys *Chem* Biol *France* **Total**	3 Phys *Chem* Biol *France* **Total**	4 or 5 Phys *Chem* Biol *France* **Total**	*Statement B*
The design of an experiment is dependent on theory about the thing that is being investigated.	60 *71* 61 *34* **57**	25 *9* 25 *14* **9**	14 *14* 13 *49* **22**	An experiment is designed to see what happens, and does not depend on theory about the thing that is being investigated.
In analysing a given data set, it is quite reasonable for different scientists to use different theoretical perspectives.	78 *83* 90 *80* **83**	9 *7* 4 *7* **7**	13 *7* 6 *9* **8**	In analysing a given data set, there is only one theoretical perspective which it is reasonable for scientists to use.
Scientists interpret data without being influenced by their theoretical assumptions.	14 *18* 19 *27* **19**	15 *11* 16 *13* **14**	70 *67* 65 *56* **65**	Scientists' theoretical assumptions influence their interpretation of data.
Scientists' ideas and theories influence their planning of data collection in experiments.	78 *72* 79 *49* **70**	13 *16* 10 *11* **12**	9 *7* 11 *37* **15**	Scientists put their ideas and theories to one side when they are planning data collection in experiments.

One data set always leads to one conclusion.	6 *9* 9 *1* **7**	23 *12* 18 *10* **16**	70 *76* 74 *84* **76**	Different conclusions can legitimately result from the same data.
Scientists plan their data analysis based on the ideas and theories that they had when designing the experiment.	85 *83* 75 *63* **77**	8 *8* 16 *17* **12**	6 *5* 8 *14* **8**	Scientists plan their data analysis without reference to the theories that they may have had when designing the experiment.
It is *not always* possible to tell which is the most powerful of two competing theories, no matter how much data are available.	68 *68* 68 *31* **64**	14 *17* 10 *10* **3**	18 *12* 23 *34* **21**	It is *always* possible to tell which is the most powerful of two competing theories if enough data are available.

The responses of both upper secondary and university students appear to be spread broadly between positions. In most cases, however, it is possible to see a clear trend towards one statement or the other. This is particularly true for university students, compared to upper secondary students. In each case, the preferred response is towards what was planned as the epistemologically more sophisticated statement, which refers to an interrelationship between theory and data in empirical work. However, in each case significant numbers of students selected the statement which was planned as epistemologically more naive.

Teachers' responses to this question were very varied. In some cases, opinion was divided (e.g. on pairs 6 and 7). In other cases, teachers overwhelmingly selected the statement written to be more epistemologically sophisticated. This indicates that some of the statements communicated as intended to teachers, whereas others did not. Certainly the teachers who responded to the survey generally appeared well able to recognise which statements were planned as the most epistemologically sophisticated.

DISCUSSION OF RESULTS

General comments

In Section 4, a number of concerns about the use of decontextualised questions for eliciting the images of science that might be drawn upon by students during labwork were raised. Data collected on this probe suggest that most upper secondary students think of data and theory as being related in designing experiments, analysing data and drawing conclusions, and that even more university students think in these terms. However, there was little internal consistency in students' responses to related pairs of statements (e.g. statements 1 and 4 relating to the design of experiments, or statements 2 and 6 about data analysis), and during pilot work where students were presented with a similar question many students selected apparently contradictory statements. This issue, and the implications for the use of decontextualised questions in probing images of science, is addressed in more detail in section 6.6.

In addition, there is some evidence for large national differences in responses to this question. For example, responses to pair 1 at the university level show between 60% and 70% physicists, chemists and biologists selecting 1 or 2, and between 13% and 14% selecting 4 or 5. However, the percentage of French univeristy students selecting these categories was 34% and 49% respectively. The problems of producing translations with the same meaning of terms like ‘design of an experiment’ and ‘theory’ was raised earlier; the broad national variations in responses to this Probe raise serious doubts as to the validity and reliability of responses.

Level-related differences
As stated previously, generally speaking more university students selected the statements planned as epistemologically sophisticated, than upper secondary students.

Subject-related differences at the University level
Generally speaking, differences between students from different subject backgrounds were small.

6.3 The Nature of Scientific Results

SAMPLE FOR THIS PROBE, AND REPORTING CATEGORIES

The sample responding to this probe is given in Table 12:

TABLE 12: SAMPLE FOR THE *NATURE OF SCIENTIFIC RESULTS* PROBE

Country	*Upper secondary*	*University physics*	*University chemistry*	*University biology*
Denmark	114	0	0	0
France	103	*	*	*
Germany	55	32	29	23
Great Britain	82	26	25	34
Greece	78	22	15	23
TOTAL	432	80	69	80

*Sample not divided by discipline. 70 students in total.

Data are reported as frequency counts, expressed as a percentage of the sample. Data from upper secondary students have not been divided into physicists, chemists and biologists. Data from University students is reported by discipline in all countries except France. Frequencies for the French university sample are reported separately.

The two parts of the probe are reported separately.

TABULATION OF DATA

Responses to the question 'My scientific statement was first put forward because:'

Students were presented with three statements of scientific knowledge publicly agreed as reliable. The statements (termed 'scientific statements') referred to the acceleration of a body subjected to a force, the chemical properties of the elements and electron configuration, and the evolution of living organisms. Students were asked to select one of these 'scientific statements'.

For the selected 'scientific statement', students had to mark 5 statements about how it was first put forward on a 1 to 5 scale.

1 - agree strongly
2 - agree [For the purpose of reporting, categories 1 and 2 were conflated.]
3 - no strong opinion
4 - disagree
5 - disagree strongly [For the purpose of reporting, categories 4 and 5 were conflated.]

Frequencies were calculated as percentages of the sample selecting one particular 'scientific statement'. Frequencies are given in Table 13, values of n being stated in the column headings:

TABLE 13: REASONS WHY THE SCIENTIFIC STATEMENT WAS FIRST PUT FORWARD

	Rel. of force and acceleration U. secondary n=151 *Uni. Phys n=60* *Uni. Chem n=18* **Uni. Biol n=14** ***Uni. France n=22*** ***UNI. TOT. n=114***			Chem. properties and electron structure U. sec. n=104 *Uni. Phys n=8* *Uni. Chem. n=34* **Uni. Biol n=10** ***Uni France n=18*** ***UNI. TOT n=70***			Evolution and mutation/ selection U. secondary n=156 *Uni. Phys n=12* *Uni. Chem. n=22* **Uni. Biol n=54** ***Uni. France n=53*** ***UNI. TOT n=111***		
	1 or 2	3	4 or 5	1 or 2	3	4 or 5	1 or 2	3	4 or 5
..as a result of data which were collected through accurate and repeated experiments	75 *72* *83* **79** ***69*** ***74***	7 *17* *11* **15** ***5*** ***13***	5 *12* *6* **7** ***27*** ***13***	64 *88* *89* **80** ***67*** ***77***	10 *0* *15* **10** ***11*** ***11***	13 *13* *6* **10** ***67*** ***10***	38 *17* *23* **44** ***70*** ***42***	15 *25* *23* **11** ***17*** ***16***	23 *58* *55* **44** ***13*** ***41***
..as a result of an imaginative proposal by a brilliant researcher	23 *45* *39* **57** ***23*** ***41***	18 *20* *33* **14** ***36*** ***25***	47 *33* *28* **29** ***41*** ***33***	29 *38* *21* **30** ***22*** ***24***	16 *25* *32* **40** ***28*** ***31***	40 *38* *47* **30** ***44*** ***43***	28 *67* *45* **52** ***30*** ***48***	19 *25* *27* **17** ***22*** ***21***	29 *8* *27* **31** ***48*** ***32***
..as a result of purely theoretical and mathematical thinking	48 *48* *50* **43** ***45***	11 *22* *17* **36** ***14***	28 *28* *33* **21** ***41***	35 *38* *29* **50** ***22***	27 *50* *29* **0** ***28***	24 *13* *41* **50** ***44***	17 *25* *27* **15** ***17***	21 *17* *23* **24** ***13***	38 *58* *45* **61** ***70***

	47	***21***	***31***	***31***	***27***	***40***	***19***	***21***	***59***
..as a result of a progressive accumulation of knowledge over a period of time	45 *52* *56* **36** ***45*** ***49***	25 *27* *28* **36** ***27*** ***28***	17 *20* *17* **29** ***27*** ***22***	63 *63* *91* **80** ***89*** ***86***	9 *13* *3* **20** ***0*** ***6***	13 *25* *6* **0** ***11*** ***9***	68 *83* *95* **79** ***100*** ***87***	4 *17* *0* **6** ***0*** ***5***	3 *0* *5* **15** ***0*** ***8***
..as a result of logical arguments and deductions	52 *68* *78* **64** ***64*** ***68***	23 *10* *22* **14** ***18*** ***14***	12 *20* *0* **21** ***18*** ***17***	58 *63* *71* **90** ***61*** ***70***	18 *0* *18* **10** ***28*** ***17***	10 *38* *12* **0** ***6*** ***11***	53 *83* *73* **80** ***100*** ***83***	12 *8* *23* **7** ***0*** ***9***	12 *8* *5* **13** ***0*** ***8***

Generally speaking, there is remarkably little difference between frequencies of responses from upper secondary students and university students.

When comparing students' responses to the three Scientific Statements, the frequencies of responses follow broadly predictable patterns. For example, more students stated that the relationship between force and acceleration was first put forward as a result of data which were collected through accurate and repeated experiments, than the theory of evolution based on mutation and natural selection. Indeed, although it is relatively easy to imagine how data about motion could be collected through accurate and repeated experiments, it is harder to envisage how this might be done in the context of evolution. Similarly, the phrase 'theoretical and mathematical thinking' does not intuitively relate to evolutionary theory in the way that it does to ideas in mechanics.

It is hard to imagine what students might have had in mind in responding to this question. For example, many students stated that the Scientific Statements were *not* put forward 'as a result of logical arguments and deductions'. It seems unlikely that such students believe that the arguments and deductions leading up to the Scientific Statement being put forward were illogical, though other interpretations do not readily come to mind. Similarly, many students stated that they did not think that their scientific statement was put forward as a result of an imaginative proposal by a brilliant researcher. Although such responses may well indicate that the students think that the Scientific Statement was put forward as a result of careful analysis of data, they may equally indicate a view that the Scientific Statements arose through work by many researchers over a sustained period.

Responses to the question 'My scientific statement is believed to be true because:'

Students then had to rate six statements about why the selected 'scientific statement' is believed to be true on the same 1 to 5 scale. Frequencies are given in Table 14, values of n being stated in the column headings:

TABLE 14: REASONS WHY THE SCIENTIFIC STATEMENT IS BELIEVED TO BE TRUE

	Rel. of force and acceleration U. secondary n=151 *Uni. Phys n=60* Uni. Chem. n=18 **Uni. Biol n=14** ***Uni. France n=22*** ***TOT n=114***			Chem. properties and electron structure U. sec n=104 *Uni. Phys n=8* Uni. Chem. n=34 **Uni. Biol n=10** ***Uni. Fr. n=18*** ***TOT n=70***			Evolution and mutation/ selection U. secondary n=156 *Uni. Phys n=12* Uni. Chem. n=22 **Uni. Biol n=54** ***Uni. France n=23*** ***TOT n=111***		
	1 or 2	3	4 or 5	1 or 2	3	4 or 5	1 or 2	3	4 or 5
..it is based on data which were collected through accurate and repeated experiments	79 *85* 89 **93** ***77*** ***85***	3 *10* 6 **0** ***5*** ***7***	5 *5* 6 **7** ***14*** ***7***	72 *88* 88 **100** ***78*** ***87***	4 *0* 12 **0** ***6*** ***7***	9 *13* 0 **0** ***17*** ***6***	46 *33* 45 **48** ***78*** ***52***	8 *25* 23 **28** ***13*** ***23***	22 *42* 32 **24** ***9*** ***24***
..it was found by a brilliant researcher.	13 *13* 22 **21** ***9*** ***15***	19 *22* 33 **14** ***36*** ***25***	56 *65* 44 **64** ***45*** ***58***	13 *13* 9 **30** ***6*** ***11***	21 *13* 24 **30** ***22*** ***23***	52 *75* 68 **40** ***72*** ***66***	12 *8* 18 **17** ***17*** ***16***	22 *25* 23 **17** ***13*** ***18***	42 *67* 59 **67** ***70*** ***66***
..it was derived from purely theoretical and mathematical thinking	43 *58* 56 **50** ***41*** ***54***	17 *22* 11 **14** ***27*** ***20***	25 *20* 33 **36** ***27*** ***25***	29 *13* 41 **20** ***39*** ***34***	18 *38* 26 **20** ***17*** ***24***	35 *38* 32 **60** ***33*** ***37***	12 *17* 32 **11** ***13*** ***16***	24 *33* 23 **20** ***22*** ***23***	39 *50* 45 **69** ***65*** ***61***

..it was the result of a progressive accumulation of knowledge over a period of time.	49 *53* *50* **57** ***45*** ***52***	26 *30* *33* **29** ***32*** ***31***	11 *17* *17* **14** ***18*** ***17***	61 *63* *82* **60** ***83*** ***77***	13 *38* *9* **40** ***6*** ***16***	12 *0* *9* **0** ***6*** ***6***	71 *92* *91* **85** ***100*** ***90***	3 *8* *5* **7** ***0*** ***5***	2 *0* *5* **7** ***0*** ***5***
..it was derived from logical arguments and deductions	59 *73* *78* **79** ***55*** ***71***	15 *17* *17* **14** ***23*** ***18***	13 *10* *6* **7** ***18*** ***11***	61 *63* *79* **100** ***83*** ***81***	16 *0* *15* **0** ***6*** ***9***	8 *38* *6* **0** ***6*** ***9***	54 *75* *86* **74** ***96*** ***81***	12 *17* *9* **15** ***4*** ***12***	10 *8* *5* **11** ***0*** ***7***
..scientists have come to agree on it, following discussion of journal articles and conference presentations	26 *27* *22* **21** ***32*** ***26***	30 *28* *39* **43** ***32*** ***32***	32 *45* *39* **36** ***32*** ***40***	32 *25* *41* **30** ***28*** ***34***	20 *13* *24* **30** ***50*** ***30***	33 *50* *35* **40** ***17*** ***33***	40 *42* *41* **52** ***43*** ***47***	17 *33* *14* **22** ***35*** ***24***	19 *25* *45* **26** ***22*** ***29***

Again, it is very difficult to interpret what many students' responses might mean, for similar reasons as described for the last part of the Probe. There is some evidence that students believe that knowledge accumulation occurs over a period of time - the majority of upper secondary and university students responded '1' or '2' to the 4th. statement. However, there is little evidence that large numbers of students think explicitly about the role of social and institutional processes in the warranting of knowledge as reliable - only about 1/3 of upper secondary and university students responded '1' or '2' to the 6th. statement.

DISCUSSION OF RESULTS

There are serious problems involved in interpreting what students' responses to this probe might mean. Teachers' responses to the question were as diverse as students. For these reasons, we are not confident that the probe yields valid or reliable information about the images of science that students might draw upon during labwork. It is, however, interesting that although many students referred to the time scale over which knowledge claims come to be believed, far fewer appear to view institutional processes such as peer review and conference presentations as important in this process.

6.4 Surprising Results

SAMPLE FOR THIS PROBE, AND REPORTING CATEGORIES

The sample responding to this probe is given in Table 15:

TABLE 15: SAMPLE FOR THE *SURPRISING RESULTS* PROBE

Country	*Upper secondary*	*University physics*	*University chemistry*	*University biology*
Denmark	104	0	0	0
France	103	*	*	*
Germany	55	32	29	23

Great Britain	82	26	25	34
Greece	78	22	15	23
TOTAL	422	80	69	80

*Sample not divided by discipline. 70 students in total.

Data are reported as frequency counts, expressed as a percentage of the sample. Data from upper secondary students have not been divided into physicists, chemists and biologists. Data from University students is reported by discipline in all countries except France. Frequencies for the French university sample are reported separately.

The two parts of the Probe are reported separately.

TABULATION OF DATA

Practical task on starch production in leaves

Students were presented with a description of results from a practical task in which school students tested whether starch had been produced in leaves kept in the light and in the dark. The results were inconclusive, as some leaves kept in both the light and the dark appeared to have produced starch, and some leaves kept in both the light and the dark appeared not to have produced any starch.

Students were then presented with six statements about what the teacher should do next. Each had to be marked 'Agree', 'Disagree' or 'Not sure'. Finally, students had to mark which they thought was the best course of action for the teacher to follow.

Students were given a space to make other suggestions in their own words. In practice, many of these comments appeared to be rewordings of statements already offered. In such cases, additional codes were not allocated. Three additional suggestions were made by a number of students:

A Explain to the class that nobody knows the answer
B Change specific factors in the design of the activity, explaining the rationale for this to the class
C Discuss the treatment of spread in biological data with students.

In practice, such responses were not all coded due to the multilingual nature of the sample and the very low frequencies at which they appeared.

Frequency counts are shown in Table 11:

TABLE 11: COURSES OF ACTION IN PRACTICAL ACTIVITY - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary Ag. Dis. ns	Uni. Phys Ag. Dis. ns	Uni. Chem. Ag. Dis. ns	Uni. Biol. Ag. Dis. ns	Uni. - France Ag Dis ns	**UNI. TOTAL**
A: Repeat, as result will become clear	52 20 24	54 23 23	44 31 25	48 21 30	37 16 41	46 23 29
B: Repeat due to faulty materials	48 24 25	56 23 22	53 25 19	60 18 20	59 11 23	57 19 21
C: Explain unexpected results as error/ anomoly	45 30 22	28 58 14	34 51 13	44 34 21	53 17 24	39 41 18
D: State accepted explanation, use it to explain unexpected results	69 12 16	77 10 13	82 6 10	81 3 14	69 11 14	77 8 13
E: Overturn accepted conclusion	3 81 12	0 90 5	4 90 5	4 93 4	1 86 9	2 90 7
F: State that standard theory is over-simplified	38 25 34	42 18 39	42 18 39	48 29 23	34 17 44	42 22 35
A preferred	20	14	18	11	7	9
B preferred	19	20	26	16	17	14
C preferred	14	5	7	13	21	7
D preferred	25	44	58	34	36	31
E preferred	0	0	0	0	0	0

F preferred	16	15	20	23	13	13

On the open response pilot to this question, two main positions about the place of data and theory in guiding the teacher's actions were stated by students in their own words. The first of these referred to collecting quality data from which an answer would logically become apparent (summarised in statements A, B, C and E). Around 50% of students at both upper secondary and university level agreed with these statements. The second main position stated in the pilot study referred to the use of theory in guiding interpretations of data (summarised by statements D and F). Between about 40 and 70% of students at each level agreed with these statements. When stating which statement was preferred, about 53% of upper secondary students and 30% of university students selected a data-focused statement, compared to 41% and 44% respectively selecting a statement involving the role of theory. A large number of university students (26%) did not, however, select a preferred statement.

It is not straightforward to interpret what students' responses to this probe may mean. For example, statement 'A' was written to convey the view that if data are collected the result will *inevitably* become clear. However, it may have been interpreted as meaning that data should help to clarify the result. Depending on which interpretation of the statement is made, responses would be differently interpreted. Having said that, it appears highly likely that large numbers of students both at the upper secondary and university levels focus upon data much more strongly than theory in interpreting results in school laboratory work. Students did not generally take this to the logical extreme and agree with statement E, presumably on the grounds that 'mistakes' and 'difficulties' in collecting data were more likely than inherent problems with theory.

Teachers' responses were remarkably consistent on statements D, E and F, and overwhelmingly emphasised the importance of drawing upon relevant theory in this situation. However, responses were more diverse on statements A, B and C. For this reason, it cannot be assumed that statements A, B and C communicated as intended to teachers and data from students ought to be treated with some caution.

Research reports into leukaemia clusters

Students were presented with a description of several studies into leukaemia clusters around chemical plants. The results were inconclusive, in that some groups reported more cases than expected around chemical plants, whereas others reported no more cases than expected.

Students were then presented with seven statements about what should be done next by the person responsible for commissioning the studies. Each had to be marked 'Agree', 'Disagree' or 'Not sure'. Finally, students had to mark which they thought was the best course of action for the future.

Students were given a space to make other suggestions in their own words.

The statements tended to be reworkings of the given statements, and no points were made by more than one or two students. Additional coding categories were not therefore allocated.
Frequency counts are given in Table 17:

TABLE 17: COURSES OF ACTION IN INVESTIGATION OF LEUKAEMIA CLUSTERS - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary Ag. Dis. ns	Uni. Phys Ag. Dis. ns	Uni. Chem. Ag. Dis. ns	Uni. Biol. Ag. Dis. ns	Uni. - France Ag Dis ns	**UNI: TOTALS**
A: Ask groups to repeat studies, to reach a firm conclusion	51 27 19	43 41 15	57 30 12	46 26 25	24 23 37	45 30 22
B: Ask other scientists to check results, to reach a firm conclusion	66 14 16	70 14 16	65 14 19	70 10 18	53 16 26	62 13 20
C: Commission more data collection: an answer will become clear	60 14 21	61 23 16	57 17 25	60 18 21	50 7 37	57 16 24
D: Decide that risk is great, close plant	17 45 35	10 56 35	5 53 39	11 45 41	23 39 31	12 48 37
E: Conclude that risk is small, allow plant to continue	11 48 37	11 63 25	17 49 31	14 51 34	6 44 46	12 52 33
F: Look for factors other than chemical plant	89 2 5	99 0 1	92 4 3	88 3 9	90 3 3	92 2 4
G: Commission theoretical work	73 7 17	85 6 9	73 8 18	81 6 11	74 1 17	78 6 14
A preferred	6	5	7	4	3	3
B preferred	13	10	13	9	8	8
C preferred	12	10	13	29	13	13
D preferred	5	5	7	3	2	2
E preferred	1	3	3	3	2	2
F preferred	41	51	67	26	31	31
G preferred	16	13	10	15	13	13

Statements A, B, C, D and E are all data focused, whereas statements F and G involve drawing upon theory to inform interpretation. Around 60% of students at each level agreed with statements A, B and C, with marginally more upper secondary students than university students. Most students (73 - 92%) agreed with statements F and G, though 17% of upper secondary students and 14% of university students explicitly disagreed with statement G (‘commission theoretical work’). Statements D and E argued that decisions about risk could be decided on the basis of available data: between 10 and 17% of students at each level agreed with these statements.

37% of upper secondary students and 28% of university students selected a data-focused statement as their preferred statement, and 57% of upper secondary students and 44% of university students selected a statement involving theory. However, 28% of university students did not select a preferred statement.

Taken together, the above evidence suggests that many students both at upper secondary and university level may well take a data-focused view in interpreting empirical data. It is perhaps not surprising that more students agreed with statements involving theory on part 2 of the probe, where the scientific knowledge claim under investigation is controversial, than agreeing with statements involving theory on part 1 where the knowledge claim in question is well established.

Teachers' responses to statements B, D, F and G were remarkably consistent, emphasising the importance of drawing upon relevant theory. However, there was some spread in responses to statements A, C and E. For this reason, it cannot be assumed that these statements communicated as intended and caution should be exercised in interpreting students' responses to these statements.

DISCUSSION OF RESULTS

General comments

The difficulty of interpreting students' responses to this closed-response format Probe have already been raised. However, taking the results together, it seems highly likely that many students in the sample at both levels are likely to focus on the data themselves, to the exclusion of underpinning theory, in interpreting findings in empirical work. As the frequencies of responses to closely related questions are quite different, it is hard to quantify how many students might take a data-focused approach. However, between 14 and 17% of students suggested that theoretical work was *not* necessary in part 2, and between 45 and 51% agreed that the groups should be asked to repeat their studies to reach a firm conclusion in part 2.

Many students on part 2 of the Probe suggested commissioning theoretical work, or looking for other factors to explain possible leukaemia clusters. By comparison, fewer students agreed that the teacher should discuss how the standard theory is oversimplified on part 1 of the Probe. It is possible that, during labwork teaching, it would be beneficial to spend more time discussing with students the complexities of using established theory to explain complex natural phenomena.

There was little evidence of subject-related differences in students' responses.

The above conclusions ought to be treated with caution, however, due to the diversity of teachers' responses to both parts of the Probe.

6.5 Interpreting Data

SAMPLE FOR THIS PROBE, AND REPORTING CATEGORIES

The sample responding to this probe is given in Table 18:

TABLE 18: SAMPLE FOR THE *SURPRISING RESULTS* PROBE

Country	*Upper secondary*	*University physics*	*University chemistry*	*University biology*
Denmark	114	0	0	0
France	103	*	*	*
Germany	55	32	29	23
Great Britain	82	26	25	34
Greece	78	22	15	23
TOTAL	432	80	69	80

*Sample not divided by discipline. 70 students in total.

Data are reported as frequency counts, expressed as a percentage of the sample. Data from upper secondary students have not been divided into physicists, chemists and biologists. Data from University students is reported by discipline in all countries except France. Frequencies for the French university sample are reported separately.

The three parts of the Probe are reported separately.

TABULATION OF DATA

Students were initially presented with background information about the potential economic importance of superconductors. A data set was then presented in the form of a plot of measured resistances of a material at temperatures in the range 50K - 100K, together with error bars. An international conference was described, where two different groups of scientists present different theoretical models of superconductivity that result in different interpretations of the data being made and different lines being drawn through the measured points. It was emphasised that the groups agreed on the form of the data, and that their differences only related to how the data could be interpreted. The first two parts of the Probe relate directly to this information.

<u>*Conflicting interpretations of the data*</u>

Students were presented with six different statements about the two interpretations of the data. They were asked to select one of the statements that best explained their own viewpoint. In addition, they were offered a space to write their viewpoint in their own words if it did not correspond with the six presented statements. In practice, however, few students suggested different viewpoints and there was little consistency in the suggestions made. Further analysis is not therefore reported.

Frequency counts are summarised in Table 19:

TABLE 19: STUDENTS' RESPONSES TO TWO DIFFERENT INTERPRETATIONS OF THE DATA - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary	University physics	University chemistry	University biology	French University	**UNI. TOTAL**
A: I agree with LIS group, as their line seems a better fit through the data points.	5	3	4	5	3	4
B: I agree with COAST group because their line seems a better fit through the data points.	5	4	3	4	3	3
C: It is unclear which group has drawn the best line. You can only decide which interpretation is better by looking at the details of the LIS and COAST models.	28	48	42	36	33	40
D: It is unclear which group has drawn the best line, but if enough data are collected it should be possible to decide between the two lines.	29	21	21	28	31	25
E: Both interpretations are acceptable. It is not possible to find out which interpretation is the best.	18	13	16	16	19	16
F: Both lines are wrong. There must be another explanation for the data.	2	0	0	1	1	1
G: Some other answer.	6	6	12	5	3	7

The closed response statements were designed to summarise the three major positions stated in their own words by students in the pilot study (see Table 3). The first of these was a data focused view, where no reference is made to underlying models when discussing data interpretation (statements A, B, D and F). 41% of upper secondary students and 33% of university students selected one of these statements. The second major position involved an interaction between underlying models and data in data interpretation, summarised by statement C. 28% of upper secondary students and 40% of university students selected this statement. The third major position was a relativist position, involving the view that different interpretations cannot be judged as better or worse than one another, summarised in statement E. 18% of upper secondary students and 16% of university students selected this statement.

Teachers' responses generally emphasised the importance of underlying models, indicating that the question communicated broadly as intended. However, a significant number of upper secondary teachers selected statement E, indicating either that the teachers held a radical relativist position or that the statement did not communicate as intended.

What should be done next?

Students were presented with eight statements of possible actions that might be taken by the scientists, presented with these conflicting interpretations of the data. For each one, students were asked to rate whether it was an appropriate course of action, not an appropriate course of action, or to state that they were not sure whether it was an appropriate course of action or not. In addition, the students were given space to suggest other courses of action that might be taken, that were not included in the eight statements. In practice, however, very few students did this and there was little consistency in what was suggested, and so no further analysis was conducted. Finally, the students were asked to state which course of action they thought would be the most important thing to do next.

Frequency counts are summarised in Table 20.

TABLE 20: STUDENTS' SUGGESTIONS ABOUT THE APPROPRIATENESS OF FUTURE ACTIONS - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary Appropriate *Not approp.* *Not sure*	University physics Appropriate *Not approp.* *Not sure*	University chemistry Appropriate *Not approp.* *Not sure*	University biology Appropriate *Not approp.* *Not sure*	French University Appropriate *Not approp.* *Not sure*	**UNIV. TOTAL** Appropriate *Not approp.* *Not sure*

	Preferred	**Preferred**	**Preferred**	**Preferred**	**Preferred**	**Preferred**
A: Draw a conclusion, based on the data avaliable, that the LIS group has explained the data correctly.	12 *59* 25 **2**	5 *78* *14* **0**	5 *72* *21* **0**	68 *71* *18* **1**	6 *63* 24 **7**	6 *71* *19* **1**
B: Draw a conclusion, based on the data available, that the COAST group has explained the data correctly.	8 *68* 24 **6**	10 *76* *10* **1**	9 *68* *21* **0**	6 *70* *21* **0**	1 *70* 21 **17**	7 *71* *18* **0**
C: Collect more data in order to prove beyond reasonable doubt which group is correct.	80 *8* 10 **28**	69 *11* *16* **19**	82 *5* *13* **20**	75 *6* *15* **33**	71 *1* 20 **21**	74 *6* *16* **25**
D: Reduce the errors in the measurements in order to prove beyond reasonable doubt that the LIS model or the COAST model gives the best interpretation.	65 *14* 19 **18**	69 *13* *10* **18**	76 *13* *11* **18**	74 *9* *15* **18**	50 *9* 34 **36**	68 *11* *19* **17**
E: It will only be possible to decide what to do next by considering the models proposed by the LIS and COAST groups.	41 *21* 35 **10**	54 *18* *25* **23**	54 *18* *28* **18**	55 *13* *30* **16**	46 *19* 27 **0**	52 *17* *27* **20**
F: Neither the LIS group nor the COAST group have explained the data correctly. The scientists need to look for other explanations.	19 *41* 38 **5**	11 *55* *30* **4**	7 *53* *39* **4**	5 *55* *36* **0**	20 *27* 44 **13**	10 *48* *37* **3**
G: Arrange for the LIS and COAST groups to meet together to decide between themselves which group has made an error.	38 *36* 24 **12**	28 *48* *23* **15**	29 *49* *22* **9**	24 *41* *33* **8**	33 *31* 29 **0**	28 *42* *26* **12**
H: The scientists should accept that there can be more than one interpretation of this data. There is no way of finding out which interpretation is the correct one.	39 *27* 32 **12**	29 *40* *28* **6**	41 *32* *28* **16**	39 *23* *36* **13**	16 *31* 44 **0**	31 *31* *34* **9**
I: The scientists should follow a different course of action.	0 **0**	0 **8**	0 **8**	1 **3**	3 **0**	1 **5**

Again, the statements were written to summarise a data-focused position (statements A, B, C, D and F), a position where data and underlying models are linked (statement E) and a relativist position (H). In addition, statement G was written to represent a view expressed in the pilot study that the LIS and COAST groups should meet to resolve the problem themselves.

Most students at both upper secondary and university levels agreed with one or more of the data focused statements, and 59% and 40% respectively selected one of the statements as their preferred statement. 41% of upper secondary students and 52% of university students agreed with statement E, 10 and 20% selecting it as the preferred statement. Interestingly, 39% of upper secondary students and 31% of university students agreed with the relativist statement (H), and 12 and 9% respectively selected this as their preferred statement. A similar number agreed with the statement that the two groups should meet to decide which group had made an error, and 12% of both school and university students selected this as their preferred response. There were some disciplinary differences in university students' responses to statements C and E, biological science students apparently being slightly more likely to select data-focused statements than physics students.

Teachers' responses to this part of the Probe overwhelmingly emphasised the importance of theory. However, a significant number of upper secondary students agreed with statement G, indicating either that the statement did not communicate as intended or that the teachers had a rather naive view of the social and institutional functioning of scientific communities.

Data on a different superconductor: what should be done next?

Finally, students were presented with a different scenario. Measurements of a similar form had been made for a different superconductor. Four possible suggestions were then made about how the data ought to be interpreted, and how relationships between the measured data ought to be drawn. Students had to select one of the four suggestions as corresponding with their viewpoint, or alternatively to state their viewpoint in their own words if it did not correspond with any of the statements. In practice, however, it was not possible to code all these responses because of problems caused by the multilingual nature of the sample. Qualitative comments are included later in this section.

Frequency counts are summarised in Table 21

TABLE 21: STUDENTS' SUGGESTIONS AS TO RELATIONSHIPS THAT MIGHT BE DRAWN BETWEEN THE MEASURED DATA POINTS - FREQUENCY COUNTS AS % OF SAMPLE

	Upper secondary	University physics	University chemistry	University biology	French University	**Uni. total**
DidaScO group: Draw a line joining each of the points. We are confident about each measurement,	4	1	1	4	6	3

so this is the best approach.						
BREM group: Consider which model could best be used to explain this data set. Once the best model has been agreed upon, a line can then be drawn through the data points.	38	43	42	39	43	42
TESME group: Use a computer to generate the best curved line through the data points. This is the best approach.	22	30	30	24	14	25
ROMA group: There is no way of knowing which is the best way to join the data points. It is up to individual scientists to make up their own minds.	21	13	13	19	20	16
I don't agree with any group! I have written my opinion below.	9	9	13	10	10	10

The DidaScO and TESME statements were written to summarise strongly data-focused positions. In practice, few students selected the DidaScO position. 26% of upper secondary students and 28% of university students selected one of these positions. It was surprising that so many students appeared to assume that computers could be used to distinguish between two different interpretations of a data set, without referring to the underlying models that might be used in programming the computer. However, it is possible that students interpreted the TESME statement as meaning that a computer would be more quick and accurate in applying a given algorithm than a person working without assistance.

The BREM statement was written to summarise a position involving both theory and data. 38% of upper secondary students and 42% of university students selected this statement.

The ROMA statement was written to summarise a radical relativist position, and was selected by 21% of upper secondary students and 16% of university students.

A number of students stating their opinion in their own words appeared to think that 2 data sets were involved in this Probe, rather than two interpretations of the same data set.

Teachers' responses to this part of the Probe were intriguing. Although most agreed with the BREM group, we were surprised to note that small numbers agreed with both the DidaScO and ROMA groups. In addition, a significant number agreed with the TESME group. This indicates either that the statements did not communicate as intended, or that the teachers deployed relativist and data-focused views in this context.

DISCUSSION OF RESULTS

General comments

Taken together, the evidence presented above seems to suggest that up to a third of upper secondary and university level science students in the sample tend to focus exclusively upon data at the expense of underlying theory, in interpreting findings from this empirical investigation. There is also evidence that between 10 and 20% of students in the sample appear to hold the radical relativist view that it is not possible to distinguish between different interpretations of data. A possible target for labwork curricula might therefore be to teach students more about the basis on which data are analysed and different models are selected: this is discussed further in section 7.

These findings should perhaps be treated with caution, as small numbers of teachers selected radical relativist statements and data-focused statements.

Level-related differences
Generally speaking, more university students than upper secondary students selected and agreed with statements mentioning explicitly links between theory and data in the interpretation of empirical work. For example, 28% of upper secondary students, compared to 40% of university students, selected statement C on part 1, 41% and 52% respectively agreed with statement D on part 2, and 38% and 42% respectively selected the BREM group’s statement on part 3. This reflected patterns of responses to other probes.

Subject-related differences at the University level
There was no strong evidence of subject-related differences in responses to this Probe. In some cases, however, biology students were slightly more likely to select data-focused responses than physics students.

6.6 The role of context

To what extent do students draw upon a similar epistemological position in a variety of different contexts, or when asked contextualised and decontextualised questions about the same epistemological issue? This is an important issue for our study: it would be useful to know whether significant numbers of students in upper secondary school and university hold stable images of science that are deployed in a variety of situations, especially if these could be identified through straightforward decontextualised questions.

In order to investigate this, individual students’ responses to probes addressing similar epistemological issues in different contexts were analysed. The data set collected contains four probes that address the relationship between knowledge claims and data, namely *Theory and Data in Scientists’ Work*, *The Nature of Scientific Results*, *Surprising Results* and *Interpreting Data*.

In sections 6.1 to 6.5, a number of clear epistemological positions were identified amongst students in the sample. For example, on *Theory and Data in Scientists’ Work* it is possible for students to articulate the position that experimental data is collected and interpreted independently from theory. Students holding this view would be expected to answer as follows on the pairs of statements:

TABLE 22:

Statement pair:	*Expected response:*
1 (design of experiments)	4 or 5
3 (interpretation of data)	1 or 2
4 (planning of experiments)	4 or 5
6 (experimental analysis)	4 or 5
7 (discrimination between theories)	4 or 5

3 students (0.4% of the sample) were identified from the whole sample, who answered with the above combination, and this was taken as evidence of a clear and consistent view that data are collected and analysed without reference to theory.

By contrast, students who viewed data collection as being closely related to theory might be expected to answer as follows on *Theory and Data in Scientists' Work*:

TABLE 23

Statement pair:	*Expected response:*
1 (design of experiments)	1 or 2
3 (interpretation of data)	4 or 5
4 (planning of experiments)	1 or 2
6 (experimental analysis)	1 or 2
7 (discrimination between theories)	1 or 2

58 students (7.9% of the sample) were identified from the whole sample, who answered with the above combination, and this was taken as evidence of a clear and consistent view that data are collected and analysed with close reference to theory.

It is interesting to note the relative sizes of these groups within the total population. An attempt was made to match pairs of statements within this Probe, so pairs 1 and 4 both refer to experimental design. There is little evidence that students' responses were consistent even within the probe.

The above students, selected on the basis of expressing consistent epistemological positions on *Theory and Data in Scientists' Work*, were treated as two subsamples of the population. The first subpopulation will be termed the *data and theory unrelated* subpopulation, whereas the second subpopulation will be termed the *data and theory related* subpopulation. Responses from the two subpopulations to the probes *Surprising Results* and *Interpreting Data* were then examined, to find out if the same epistemological positions were used. Only those parts of probes that required students to articulate a strong position that data and theory are unrelated, or related, were used.

[The two subpopulations' responses on *The nature of Scientific Results* were not compared, due to concerns about this question expressed earlier.]

COMPARISON OF SUBPOPULATION RESPONSES ON *THEORY AND DATA IN SCIENTISTS' WORK* AND *SURPRISING RESULTS*

Members of the two subpopulations might be expected to make the following statements on the *Surprising Results* probe:

TABLE 24

Part of probe:	*Expected response: 'data and theory unrelated'*	*Expected response: 'data and theory related'*
Case 1 statement A (repeat the experiment to collect more data)	Agree or Not sure	Disagree or Not sure
Case 1 statement B (repeat the experiment but with different plants)	Agree or Not sure	Disagree or Not sure
Case 2 statement C (collect more data of the same kind)		Agree or Not sure

It is acknowledged that the closed-response statements on this probe might be interpreted in a variety of ways, and that students might not therefore make the intended interpretation of the statements.

3 students from each subpopulation answered consistently across the two probes. This is 100% of the *theory and data unrelated* subpopulation (which contained 3 students only), and 5% of the *theory and data related* subpopulation (58 students). The 6 students answering consistently represents 0.82% of the total sample.

COMPARISON OF SUBPOPULATION RESPONSES ON *THEORY AND DATA IN SCIENTISTS' WORK* AND *INTERPRETING DATA*

Members of the two subpopulations might be expected to make the following statements on the *Surprising Results* probe:

TABLE 25

Part3 of question	*Expected response:*
Theory and Data Unrelated subpopulation	Agreement with DidaScO group or TESME group
Theory and Data Related subpopulation	Agreement with BREM group

2 students from the *theory and data unrelated* subpopulation (66% of the subpopulation and 0.27% of the total sample) answered consistently on the two probes, and 13 students from the *theory and data related* subpopulation answered consistently (22% of the subpopulation and 1.76% of the total population).

COMPARISON OF SUBPOPULATION RESPONSES ON ALL THREE PROBES (*THEORY AND DATA IN SCIENTISTS' WORK, SURPRISING RESULTS* AND *INTERPRETING DATA*)

The number of students from the *data and theory unrelated* subpopulation answering consistently across the three probes are shown in the following table:

TABLE 26

Number of students in data and theory unrelated subpopulation, as identified on the Theory and	*Number of students in subpopulation answering consistently on Surprising*	*Number of students in subpopulation answering consistently on Surprising*

Data in Scientists' Work probe:	*Results probe:*	*Results and Interpreting Data probes:*
3 (0.41% of total sample)	3 (0.41% of total sample)	2 (0.27% of total sample)

The number of students from the *data and theory related* subpopulation answering consistently across the three probes are shown in the following table:

TABLE 27

Number of students in data and theory related subpopulation, as identified on the Theory and Data in Scientists' Work probe:	*Number of students in subpopulation answering consistently on Surprising Results probe:*	*Number of students in subpopulation answering consistently on Surprising Results and Interpreting Data probes:*
58 (7.86% of total sample)	3 (0.41% of total sample)	0

Taken together, these data do not provide convincing evidence that individual students' responses to decontextualised questions can be used to predict how they will respond to situations placed in specific contexts. This finding is in line with findings from previous work, which has indicated that individuals draw upon a range or profile of images of science in making sense of different examples and situations (Driver et al., 1996). However, the extent to which findings from contextualised survey questions can be used to predict an individual's actions during labwork remains open to question, and is discussed in the next section.

7 Discussion: the significance of findings from the study for labwork

7.1 Summary of findings

Throughout this report, attention has been drawn to the methodological difficulty of probing students' images of science, and specific reservations have been voiced about each of the five Probes used in this study. However, it is interesting to note that the population of students made very similar responses across a number of the Probes. In addition, Case Study evidence suggests that similar patterns of reasoning were being used by students engaged in labwork (Psillos et al., Working Paper 7). Taken together, this indicates that significant numbers of students at the upper secondary and university levels are likely to draw upon the images of science reported in this study when working with knowledge claims and data during empirical enquiry.

In section 1, a case was made for viewing individual students as drawing upon a range of images of science during their various activities, different images of science being used in different situations. During empirical investigations, we would expect an individual student to draw upon a range of different images of science when presented with different activities. For example, we would expect many students to suggest different courses of action if experimental data were not consistent with a canonical knowledge claim (e.g. green leaves produce starch in sunlight, using water and carbon dioxide), than in cases where experimental data are inconclusive with respect to a more controversial knowledge claim (e.g. there are leukaemia clusters around particular chemical plants). Bearing this view in mind, it is not appropriate to use findings from either the contextualised or decontextualised Probes used in this study to state that 'X% of students reason in such-and-such a way': we would expect the numbers of students reasoning in particular ways to vary considerably according to the context in hand. Findings from the study are therefore used to indicate images of science drawn upon by *significant numbers of students in the population*, across a range of specified contexts, rather than the images of science likely to be drawn upon by *any individual student in any specified context*.

The images of science drawn upon by students in the sample are summarised below around the Research Questions which were investigated in the study (see Sections 2 and 4).

* *RQ1: Do students see measured data as a 'perfect' copy of reality, or do they view measured data as being subject to some uncertainty? What do they see as the sources of uncertainty in measured data? How do they overcome these uncertainties and select a value? Do they recognise the difference between <u>accuracy</u> and <u>precision</u>?*

This question was investigated through two Probes, *Differences in Values from Measurement* and *Surprising Results*. In both probes, there was evidence that many students (between 30% and 60%) appeared to think that with good enough apparatus and enough care it is possible to make a perfect measurement of a quantity (e.g. the mass of a sample of oil, whether leaves on plants in sunlight were producing starch, whether leukaemia clusters exist near a chemical plant). Indeed, around 5% of students stated explicitly that 'real' values ought to correspond to measured data. These viewpoints resulted in students explaining all spread and variation in data in terms of imperfections in technique and mistakes by the scientists collecting data.

In practice, virtually all students on the *Differences in Values from Measurement* Probe were aware that algorithms are generally used to generate a value from a set of measured data. Generally, students suggested using the arithmetic mean. However, many students' responses suggested a lack of explicit understanding of the basis of estimating values from data sets. For example, around 30% of students suggested that equal confidence could be ascribed to two estimates of a value, based on different data sets with different spread, as the value of the mean in each case was the same.

Around half the students in the sample selected responses indicating some understanding that the spread of a set of measurements has implications for the confidence that can be ascribed to estimates calculated from the set. Approximately 10% more university students than upper secondary students selected responses of this kind.

* *RQ2: Do students believe that the only way to judge the quality of a measurement is from a known 'true' result, or do they believe that the quality of a measurement can be judged from a set of repeated measurements? If so, do they distinguish the <u>accuracy</u> and <u>precision</u> of a measurement?*

This research question was probed by the *Differences in Values from Measurement* Probe. As stated for RQ1, many students appeared not to appreciate the significance for an estimated value of the spread of data from which the estimate was made. In practice, the closed response questions used in this study were not particularly useful in determining whether students distinguished the accuracy and precision of measurements. However, as stated for RQ1, small numbers of students stated explicitly that estimates of quantities ought to correspond to measured data, and many students seemed to judge the quality of estimates by comparing the values of estimates themselves, rather than the data from which the estimates were made.

* *RQ3: When working with data sets, do students see procedures like joining data points with lines of 'best fit' or smooth curves as routine heuristics, or alternatively as a process of proposing tentative hypotheses?*

This research question was probed by the *Interpreting Data* Probe. Three major views were apparent. Some students (up to 80% in some cases) appeared to focus their attention entirely upon data, assuming that enough data of sufficient quality would show relationships unproblematically and that collecting data of sufficient quality was a straightforward process. Students who suggested that the use of computers in data handling would solve problems in data interpretation were interpreted as holding such a view.

Other students (typically around 30%) appeared to focus on the relationship between data and underlying models, stating that the quality of data can only be judged in terms of the models, and that conclusions about relationships can therefore only be made if due consideration is given to the underlying model.

Other students (between 20% and 40%) appeared to take a radical relativist view, arguing that there is no way that one relationship between quantities can be judged as superior to another, and that it is completely acceptable for different scientists to hold different opinions.

In many cases both university and upper secondary students were equally likely to draw upon the three viewpoints described above, though in other cases around 20% more university students referred to relationships between data and models.

* *RQ5: Do students think that scientific theories 'emerge' from data, or do they think of scientific theories and data as being related in a more complex way? If so, how do they think that scientific theories and data are related? In particular, do they think that a given experiment is open to more than one interpretation?*

A number of the Probes provided evidence relating to this research question (*Theory and Data in Scientists' Work, The Nature of Scientific Results, Surprising Results, Interpreting Data*). Drawing upon these Probes, three broad viewpoints are apparent, not unsimilar to those described for RQ3.

Many students (often as many as 50%) appeared to think that most or all questions about natural phenomena are answerable by collecting observational data and looking for correlations, and that explanatory theories emerge from this data in a logical way as there is only one logical explanation. For example, they selected responses stating that the design and interpretation of experiments is independent of theory, or that results will always become clear if enough data of sufficient quality are collected.

Other students appeared to think that theory is involved in making decisions about data collection, and the evaluation of data (between 30% and 60%, depending on the context).

Amongst these students, some selected statements implying that there is only one logical interpretation of a given data set (around 15%) whereas others selected statements implying that there is more than one legitimate interpretation (around 70%).

Other students appeared to hold a radical relativist position, selecting statements implying that there is no rational basis for evaluating knowledge claims and data (typically around 20%).

* *RQ9: Do students recognise that different courses of action are appropriate in scientific investigations depending on the status of the scientific knowledge claim under investigation?*

This research question was probed by the *Surprising Results* Probe. Unsurprisingly, virtually no students in the sample suggested that canonical scientific knowledge such as starch production in plant leaves in sunlight should be overturned on the basis of findings from student labwork. When selecting statements relating to student labwork on starch production in leaves, the focus of many students' choices was upon why the data did not match a known outcome.

On the other hand, many more students (around 90% for both upper secondary and university students) referred to extending theoretical understanding when discussing the possibility of leukaemia clusters around a chemical plant, which was taken as evidence of some recognition that the knowledge claim in question was of more problematic status within the scientific community than starch production in leaves. However, between 11% and 17% of students agreed with a statement that a definite decision could be made on the given, inconclusive data.

This appears to suggest that, with the contexts used in this study, most students in the population were well able to recognise that some knowledge claims enjoy more support in the scientific community than others, and are therefore less likely to be overturned or modified on the basis of relatively small amounts of empirical investigation.

7.2 Recommendations for labwork teaching arising from the study

In Section 1 we suggested that findings from this study could be used in identifying possible curricular goals for labwork at the upper secondary and university levels, and in identifying teaching approaches in labwork at these levels. In Section 7.1, evidence was presented that large numbers of students in the sample drew upon images of science that are potentially unhelpful during labwork learning. We therefore make the following recommendations about curricula and teaching approaches as they affect student learning through labwork in the upper secondary and university curriculum:

1 Many students are likely not to recognise the epistemological basis of routine algorithmic procedures used for data handling during labwork, such as estimating values from sets of data and drawing lines and curves through measured data points. In some cases, this is likely to lead to students taking inappropriate actions during their labwork learning (such as assuming that computers can solve problems of data analysis, not recognising the need for scientists to programme computers how to handle data according to specific requirements). Findings from this study suggest that individual students draw from a range of images of science in acting in various situations. For many students, it may therefore be necessary to introduce ideas about the epistemological basis of routine algorithms for data analysis, as well as to give students experience and practice at applying this reasoning in a variety of appropriate labwork contexts.

We therefore recommend that explicit teaching is planned in the curriculum to teach students about the epistemological basis of data analysis in contexts close to those likely to be encountered during routine labwork.

We further recommend that the epistemological basis of data handling is made explicit, rather than treated as understood, during labwork teaching activities involving the use of such algorithmic procedures.

2 Many students are likely to see knowledge claims as emerging directly from the logical analysis of data, not recognising how particular theories and models help to shape scientists' ways of evaluating and interpreting data. This may lead to inappropriate behaviour during labwork, such as students not recognising how theory might be drawn upon during experimental design, analysis and interpretation. Findings from this study suggest that many students are likely to focus upon data alone, and that some students will adopt a radical relativist position, that there is no rational basis for evaluating knowledge claims during empirical work.

We therefore recommend that, through the science curriculum, students are exposed to philosophical ideas about relationships between knowledge claims and data. These ideas should be presented to students in contexts close to those likely to be encountered during other aspects of their studies in science.

We further recommend that students' attention be drawn explicitly to the role of theory in the design, interpretation and analysis of data during a range of labwork activities.

3 Some students appear likely to draw strong conclusions from empirical investigations, based on inconclusive evidence.

We therefore recommend that, through the science curriculum, students are exposed to examples of empirical investigations and issues related to drawing conclusions from those investigations. Examples might include case studies from the work of professional scientists, or students' own labwork.

Acknowledgement
The work reported in this paper was supported by the EC project *Labwork in Science Education.* The work of Claire Middleton on coding and analysis of data is gratefully acknowledged.

References

Aikenhead, G., Fleming, R. and Ryan, A. (1987) *High-school graduates' beliefs about science-technology-society. 1. Methods and issues in monitoring student views.* Science Education, 71, (2): 145 - 161.

Bandiera, M., Tarsitani, C., Torracca, E. and Vicentini, M. (1998) *Working Paper 4, LSE Project.*

Carey, S., Evans, R., Honda, M., Jay, E. and Ungar, C. (1989) *'An experiment is when you try it and see if it works': a study of grade 7 students' understanding of the construction of scientific knowledge.* International Journal of Science Education, 11, (5), 514 - 529.

Driver, R., Leach, J., Millar, R. and Scott, P. (1996) *Young people's images of science*. Buckingham: Open University Press.

Kuhn, D., Amsel, E. and O'Loughlin, M. (1988) *The development of scientific thinking skills.* London: Academic Press.

Larochelle, M. and Désautels, J. (1991) *'Of course, it's just obvious': Adolescents' ideas of scientific knowledge*. International Journal of Science Education, 13 (4), 373-389.

Leach, J. T., Driver, R. H., Scott, P. H., Wood-Robinson, C. (1996): *Children's ideas about ecology 3: ideas found in children aged 5 - 16 about the interdependency of organisms*. International Journal of Science Education, 18, (2), 129-142.

Lubben, F. and Millar, R. (1996) *Children's ideas about the reliability of experimental data*. International Journal of Science Education, 18, (8), 955-968.

Millar, R., Le Maréchal, J-F. and Tiberghien, A. (1998) *A map of the variety of labwork in Europe.* Working Paper 1, LSE project.

Miller, J. (1994) *The relationship between biomedical understanding and public policy.* Paper presented at the 1994 international conference on the public understanding of science, London.

Niedderer, H., Bethge, T., Meyling, H. and Schecker, H. (1992) *Epistemological beliefs of students in high school physics.* Paper presented at the annual meeting of NARST, Boston, March 1992.

Nott, M. and Wellington, J. (1993) *Your nature of science profile: an activity for science teachers.* School Science Review, 75, (270), 109 - 112.

Nott, M. and Wellington, J. (1997) *Probing teachers' views of the nature of science: How should we do it and where should we be looking?* in G. Welford, J. Osborne and P. Scott (Editors) Research in science education in Europe: Current issues and themes pp 283-294.

Psillos, D. and Niedderer, H. (1997) *Working Paper 7, LSE Project.*

Ryder, R., Leach, J. and Driver, R. (in press): *Undergraduate science students' images of science.* Journal of Research in Science Teaching.

Ryder, J., Leach, J. and Driver, R. (1997) *The interaction between undergraduate science students' images of the nature of science and their experiences of learning science.* Paper presented at the first conference of the European Science Education Research Association, Rome, 2-6 September 1997.

Samarapungavan, A. (1992) *Scientists' conceptions of science: a study of epistemic beliefs.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1992.

Séré, M-G, Journeaux, R. and Larcher, C. (1993) *Learning the statistical analysis of measurement errors.* International Journal of Science Education, 15 (4), 427-438.

Tamir, P. and Zohar, A. (1991) *Anthropomorphism and teleology in reasoning about biological phenomena*. Science Education 75 (1) 57-68.

Appendix 1: The full text of the probes

Differences in Values from Measurement

Two groups of nutritionists have been asked to measure the mass of 100 cm^3 of nut oil. Each group takes nine samples of 100 cm^3 of the oil from a large container and weighs each sample. These are their results, after having sorted them into ascending order:

Measurements in grams:

Group A	81.9	83.5	86.5	87.1	87.3	87.5	87.5	90.5	92.1	(average 87.1)
Group B	84.9	85.7	86.6	86.9	87.0	87.3	88.2	88.5	88.8	(average 87.1)

What should Group A state as their result for the mass of 100 cm^3 of nut oil? *Please fill in your answer in the box below:*

Please explain your reasoning:

With which of the following statements do you most closely agree?

A	We can be more confident in Group A's result, because the range between the largest and the smallest measurement is greater.	
B	We can be more confident in Group A's result, because two of their measurements agree.	
C	We can be more confident in Group A's result, because one of their measurements is the same as their average.	
D	We can be more confident in Group B's result, because the range between the largest and the smallest measurement is less.	
E	We can be equally confident in either Group's result, because both sets of measurements have the same average.	
F	We cannot decide which Group's result we can use with greater confidence.	

Another two groups of nutritionists have been asked to measure the mass of 100 cm^3 of soya oil. Each group takes nine samples of 100 cm^3 of the oil from a large container and weighs each sample. These are their results:

Measurements in grams:

Group A	82.3	82.6	83.0	83.2	83.3	83.8	83.9	84.2	84.3	(average 83.4)
Group B	81.5	82.2	83.1	83.8	84.5	84.6	84.7	85.9	86.6	(average 84.1)

Looking at both these sets of results, what would you state as the mass of 100 cm^3 of soya oil? *Please fill in your answer in the box below:*

Please explain your reasoning:

With which of the following statements do you most closely agree?

A	We can be more confident in Group A's result, because the range between the largest and the smallest measurement is less.	
B	We can be more confident in Group B's result, because the range between the largest and the smallest measurement is greater.	
C	We can be equally confident in either Group's result, because the value 83.8 is in both sets of measurements.	
D	We can be equally confident in either Group's result, because there is not much difference between the averages.	
E	We cannot decide which Group's result we can use with greater confidence.	

A group of nutritionists wants to compare the masses of sunflower oil and olive oil. They make nine measurements on 100 cm^3 samples of each sort of oil. These are their results:

Measurements in grams:

Sunflower oil	92.1	92.4	92.6	92.7	93.1	94.0	94.1	94.3	94.4	(average 93.3)
Olive oil	92.5	92.9	93.0	93.2	93.4	93.7	94.1	94.2	94.5	(average 93.5)

With which of the following statements do you most closely agree?

A	The mass is greater for olive oil than for sunflower oil, because the average is larger.	
B	There is no difference in the masses, because the range of measurements in each set is much bigger than the difference between the averages.	
C	There is no difference in the masses, because the value 94.1 is in both sets of measurements.	
D	We cannot be sure that there is a difference in the masses, because the range of measurements in each set is much bigger than the difference between the averages.	
E	We cannot be sure that there is a difference in the masses, because the value 94.1 is in both sets of measurements.	

Theory and Data in Scientists' Work

The following pairs of statements are about how scientists work with theory and data. In each case, please mark your opinion on the scale provided.

1 indicates a strong agreement with Statement A, or that statement A is usually true. 5 indicates a strong agreement with Statement B, or that statement B is usually true.
2 indicates agreement with Statement A, or that statement A is generally ture. 4 indicates agreement with Statement B, or that statement B is generally true.
3 indicates that you have no strong opinions either way, or that you think both statements have some merit.

Statement A	1	2	3	4	5	*Statement B*
The design of an experiment is dependent on theory about the thing that is being investigated.						An experiment is designed to see what happens, and does not depend on theory about the thing that is being investigated.
In analysing a given data set, it is quite reasonable for different scientists to use different theoretical perspectives.						In analysing a given data set, there is only one theoretical perspective which it is reasonable for scientists to use.
Scientists interpret data without being influenced by their theoretical assumptions.						Scientists' theoretical assumptions influence their interpretation of data.
Scientists' ideas and theories influence their planning of data collection in experiments.						Scientists put their ideas and theories to one side when they are planning data collection in experiments.
One data set always leads to one conclusion.						Different conclusions can legitimately result from the same data.
Scientists plan their data analysis based on the ideas and theories that they had when designing the experiment.						Scientists plan their data analysis without reference to the theories that they may have had when designing the experiment.
It is *not always* possible to tell which is the most powerful of two competing theories, no matter how much data are available.						It is *always* possible to tell which is the most powerful of two competing theories if enough data are available.

The nature of scientific results

Read the following statements:

•The acceleration of a body subjected to a force is proportional to the force itself.	
•Chemical properties of the elements can be explained by the configuration of electrons in the element’s atoms.	
•The evolution of living organisms is due to random mutations and selection.	

Each of these statements is generally believed by scientists, and so we will call them SCIENTIFIC STATEMENTS.

Choose one of these SCIENTIFIC STATEMENTS and keep it in mind as you answer the questions below. Mark the SCIENTIFIC STATEMENT you have chosen.

In the following table, you will find a number of different explanations as to how your chosen SCIENTIFIC STATEMENT was first put forward. For each explanation, please mark your agreement/disagreement in the space provided.

1: Agree strongly
2: Agree
3: No strong opinion
4: Disagree
5: Disagree strongly

My SCIENTIFIC STATEMENT was first put forward:	1	2	3	4	5
.. as a result of data which were collected through accurate and repeated experiments.					
.. as a result of an imaginative proposal by a brilliant researcher.					
.. as a result of purely theoretical and mathematical thinking.					
.. as a result of a progressive accumulation of knowledge over a period of time.					
.. as a result of logical arguments and deductions.					

In the next table, some ideas are presented as to why your SCIENTIFIC STATEMENT is believed to be true. Please mark your opinion about each one in the space provided:

My SCIENTIFIC STATEMENT is believed to be true because:	1	2	3	4	5
.. it is based on data which were collected through accurate and repeated experiments.					
.. it was found by a brilliant researcher.					
.. it was derived from purely theoretical and mathematical thinking.					
.. it was the result of a progressive accumulation of knowledge over a period of time.					
.. it was derived from logical arguments and deductions.					
.. scientists have come to agree on it, following discussion of journal articles and conference presentations.					

Surprising Results

Case 1

A class is carrying out a practical task in the laboratory. They are testing samples of leaves from plants which have been kept in the dark for several days, and from other plants which have been kept constantly in the light. The students are testing the samples of leaves for the presence of starch. The textbooks say that there should be starch in the leaves from plants kept in the light, but no starch in the leaves of plants kept in the dark.

Some groups of students in the class get the expected results. However, some groups of students get negative starch test results for all of their leaves, whether kept in the light or dark. And some groups get a positive starch test for all of their leaves - both those that have been kept in the light *and* those that have been kept in the dark.

If you were the teacher in charge of this class, what would you do now? *For each of the following statements tick one box to indicate whether you agree, disagree, or are not sure.*

		Agree	Disagree	Not Sure
A	Repeat the experiment to check the results: if the students work carefully enough, the expected result will become clear.			
B	Repeat the experiment, but with different plants and test materials in case the first ones were faulty in some way. If enough data are collected, the expected result will become clear.			
C	Draw attention to the results of the groups who got the expected answer, and explain the other groups' results in terms of experimental errors and anomolies.			
D	Tell the students the accepted explanation, and also use it to explain how the unexpected results might have come about.			
E	Tell students that the conclusion from the experiment is that the standard theory is incorrect and that they should learn the result their experiments indicate.			
F	Tell students that the conclusion from the experiment is that the standard theory is over-simplified, and the connection between light and starch in leaves is more complex.			

If you were the teacher, which ONE of the above courses of action do you think would be best? *Please write a letter in the space provided*

If you have any other suggestions about what the teacher should do, please explain them here:

Case 2

A small number of children living in one area of the country are found to be suffering from leukaemia. The number of cases is small, but more than you would normally expect in a population of that size. It is suggested that a particular substance, which is emitted in small quantities from a chemical plant in the area, may be the cause. The chemical plant, however, is a major source of employment and its closure would put many people out of work.

Several groups of scientists were commissioned to investigate the number of cases of leukaemia in the areas around the chemical plant. For comparison, other groups of scientists were commissioned to look at the incidence of leukaemia in similar communities which are far away from any chemical plant.

Some groups report finding more cases of leukaemia than expected in the region near a chemical plant. Some other groups report that the number of cases of leukaemia is no higher than you would expect in the general population. Some groups report that there are other areas of the country with a slightly higher than expected incidence of leukaemia, but with no chemical plant nearby.

If you were the person responsible for commissioning the reports, what would you do now? *For each of the following statements tick one box to indicate whether you agree, disagree, or are not sure.*

		Agree	Disagree	Not Sure
A	Ask the same groups to repeat their studies, in order to check their results and reach a firm conclusion.			
B	Ask other scientists to check the results and analysis of the original researchers, in order to reach a firm conclusion.			
C	Commission further studies of the same kind: if enough data are collected, an answer should become clear.			
D	Decide that people's safety is at risk, and recommend that the plant should be closed.			
E	Decide that the risk to people around the chemical plant is very small, and recommend that the plant be allowed to continue operating.			
F	Commission further scientific work, to identify factors *other* than the chemical plant that might be responsible for causing leukaemia.			
G	Commission further scientific work, of a more theoretical kind, to try to find an explanation of *how* particular emissions might cause leukaemia.			

Which ONE of the above courses of action do you think would be best? *Please write a letter in the space provided*

If you have any other suggestions about what the person responsible for commissioning the reports should do, please explain them here:

Interpreting Data

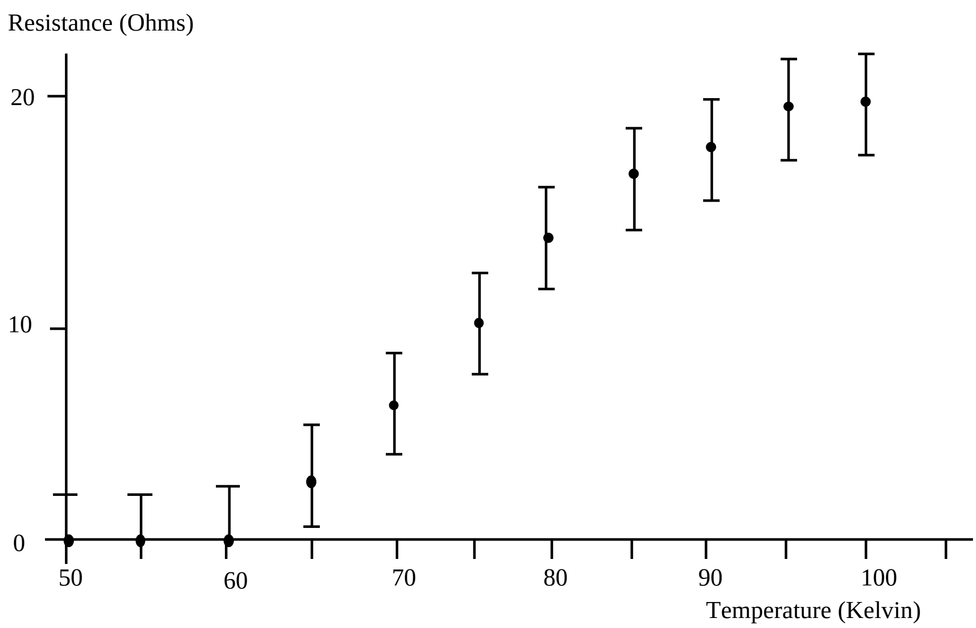
This question is about a real issue being worked on by scientists at the moment. In order to answer this question, you *do not* need to understand the technical details of the scientists work: the main focus of the question is upon the ways in which scientific data is analysed and interpreted.

The first two pages of this question present some background information about the scientists' work. Please read this carefully before answering the questions on the third page.

Part 1

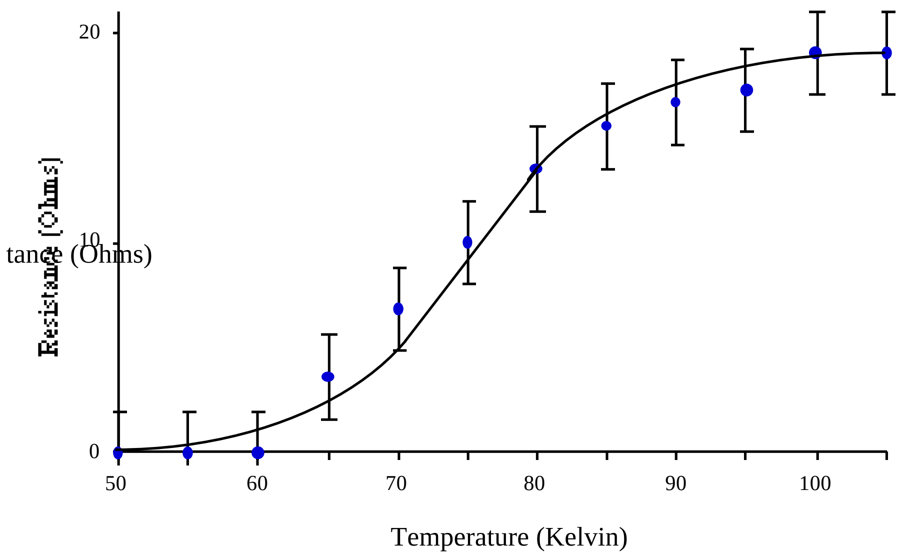
Superconductors are very special materials. Using superconductors to transfer electricity around the country could lead to a significant reduction in energy consumption. This is because these materials have zero resistance to electricity at low temperatures.

A group of scientists from around the world are investigating the properties of a new superconductor which could be used to carry electricity. One of these groups - the LIS group - has made measurements of the electrical resistance of this superconductor as the temperature is changed. The LIS experiment was performed under carefully controlled conditions, and other scientists have been able to repeat the measurements in their own laboratories and get virtually identical results. The LIS results, with error bars, are given below:



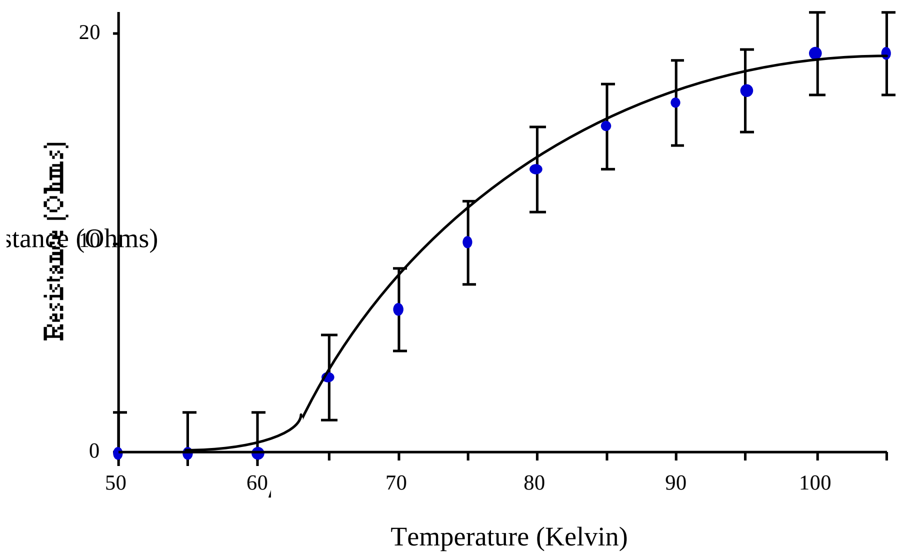
At an international conference the LIS group meets with several other research groups to discuss the analysis and interpretation of this data. The LIS group have their own theoretical model of superconductivity to explain the fall in electrical resistance. The LIS model leads to an interpretation of the data that is shown by the line on the diagram below:

LIS GROUP INTERPRETATION



The COAST group have developed a *different* theoretical model of superconductivity. The COAST model leads to an interpretation of the data that is shown by the line on the diagram below:

COAST GROUP INTERPRETATION



What do you think about these two interpretations of the data?

Please tick ***one*** *of the boxes below.*

A	I agree with the LIS group because their line seems a better fit through the data points.	
B	I agree with the COAST group because their line seems a better fit through the data points.	
C	It is unclear which group has drawn the best line. You can only decide which interpretation is better by looking at the details of the LIS and COAST models.	
D	It is unclear which group has drawn the best line, but if enough data are collected it should be possible to decide between the two lines.	
E	Both interpretations are acceptable. It is not possible to find out which interpretation is the best.	
F	Both lines are wrong. There must be another explanation for the data.	
G	Some other answer. *Please explain below:*	

Part 2

The scientists assembled at the conference now have to decide what to do next. Below, a number of suggestions are listed.

For each suggestion, please say whether you think that it is an appropriate course of action, not an appropriate course of action, or whether you are not sure whether it is an appropriate course of action or not.

		Appropriate	Not appropriate	Not sure
A	Draw a conclusion, based on the data available, that the LIS group has explained the data correctly.			
B	Draw a conclusion, based on the data available, that the COAST group has explained the data correctly.			
C	Collect more data in order to prove beyond reasonable doubt which group is correct.			
D	Reduce the errors in the measurements in order to prove beyond reasonable doubt that the LIS model or the COAST model gives the best interpretation.			
E	It will only be possible to decide what to do next by considering the models proposed by the LIS and COAST groups.			
F	Neither the LIS group nor the COAST group have explained the data correctly. The scientists need to look for other explanations.			
G	Arrange for the LIS and COAST groups to meet together to decide between themselves which group has made an error.			
H	The scientists should accept that there can be more than one interpretation of this data. There is no way of finding out which interpretation is the correct one.			
I	The scientists should follow a different course of action. *Please explain below:*			

Which ONE of the above courses of action do you think would be the most important thing to do next? *Please write a letter in the space provided:*

Part 3

The research groups at the conference now consider some measurements of the change in electrical resistance with temperature for a *different* superconductor. Once each data point is plotted onto a graph, with error bars, several different suggestions are made about what should be done next:

DidaScO Group
Draw a line joining each of the points. We are confident about each measurement, so this is the best approach

BREM Group
Consider which model could best be used to explain this data set. Once the best model has been agreed upon, a line can then be drawn through the data points.

TESME Group
Use a computer to generate the best curved line through the data points. This is the best approach.

ROMA Group
There is no way of knowing which is the best way to join the data points. It is up to individual scientists to make up their own minds.

In the table below please put a tick to show which group, if any, you agree with, and explain your reasoning.

I agree with the DidaScO group	
I agree with the BREM group	
I agree with the TESME group	
I agree with the ROMA group	
I don't agree with any group! I have written my opinion below.	

Please explain your reasoning below:

....

....

....

....

....

……

…………………………………