

Concepts of evidence and their role in open-ended practical investigations and scientific literacy; background to published papers.

Richard Gott and Ros Roberts
The School of Education, Durham University, UK.

December 8, 2008

Acknowledgements We wish to thank Judith Glaesser whose work on the QCA analysis has informed this report.

Contents

1	Introduction	6
1.1	The structure of the reporting of the research	6
1.2	The historical background	6
1.3	Methodological issues	7
2	Background: the procedural component of a science curriculum	8
2.1	Introduction	8
2.2	What understanding is important?	8
2.3	Other perspectives on the procedural component	9
2.3.1	Piagetian schema	9
2.3.2	Inquiry	10
2.3.3	Nature of Science	10
2.4	The Procedural component of the curriculum	10
2.4.1	The skills perspective	10
2.4.2	An understanding of ideas about evidence perspective	11
2.5	Teaching the procedural component	13
2.5.1	Teaching from the skills perspective (cells 1 and 2)	14
2.5.2	Procedural component separately, (cell 1)	14
2.5.3	Procedural and substantive components together, (cell 2)	14
2.5.4	Understanding of evidence perspective (cells 3 and 4)	14
2.5.5	Procedural component separately, (cell 3)	15
2.5.6	Procedural and substantive components together, (cell 4)	16
2.6	Assessing the procedural component	16
2.6.1	Assessment from the skills perspective	16
2.6.2	Assessment from an ideas about evidence perspective	17
2.7	How these ideas inform our research	17
2.8	A simplified typology of tasks	19
2.8.1	Type 1: Low on substantive understanding (Knowledge-lean)	19
2.8.2	Type 2: Substantive understanding plays a helpful part	19
2.8.3	Type 3: Substantive understanding is required	19
3	The theoretical framework	21
3.1	Introduction	21
3.2	Models as descriptions and tools for analysis	21
3.3	Defining terms	21
3.4	The Concepts of Evidence: model 1	22
3.5	Evidence in investigations	24
3.5.1	A skills approach to investigative work	24
3.5.2	An ‘understanding’ approach	24
3.5.3	The role of substantive ideas	25
3.5.4	Working iteratively	27
3.6	Toulmin and argumentation	27
3.6.1	Evidence	30
3.7	Looking forward and Looking back	30

3.8	Concepts of evidence - investigations and claims	33
3.9	Putting the models together	35
3.10	An example	36
3.11	The theoretical structure as applied to the research	38
3.12	Some additional comments	38
3.12.1	Under-specification	38
3.12.2	What we do not include in our structure	39
4	Methodology	40
4.1	The sample	40
4.2	The instruments	40
4.2.1	Biographical information	40
4.2.2	The evidence pre-test and post-test	42
4.2.3	The open-ended investigation - cohort 2004	42
4.2.4	The open-ended investigation - cohort 2005	43
4.2.5	The transfer task	44
4.3	The sample described	45
4.3.1	Cohorts 04 and 05	45
4.3.2	Social class and socio-economic status	46
4.3.3	The sub-samples used in the analysis	48
4.4	A note about the language of necessity and sufficiency	49
4.4.1	Sets and logic	49
5	An outline of the teaching programme	51
5.1	The BAEd course at Durham	51
5.2	The modules targeted at substantive understanding	51
5.3	The Year 2 evidence module	51
5.4	The rationale behind the teaching	52
5.4.1	The logical sequence	52
5.5	The evidence module	52
5.5.1	Examples of teaching intervention	55
5.6	A 'task analysis' using observations from teaching	56
5.6.1	Introduction	56
5.6.2	The Written evidence test	57
5.6.3	The investigation	57
5.6.4	The Issue task	59
5.6.5	What issues does this raise for the research and its explanatory framework?	60
5.6.6	Some tentative hypotheses derived from teaching experience	61
5.7	Research questions	61
A	The evidence test	62
A.1	Coding of the test	62
A.2	Tests for normality and item statistics	62
A.2.1	Normality	62
A.2.2	Various statistics related to the evidence pre and post-tests	62
	Bibliography	68

List of Figures

2.1	Bullseye	13
3.1	Model 1: the bullseye	23
3.2	A skills approach within an explanatory focus	24
3.3	Model 2: An ‘ideas’ approach within a problem-solving focus	25
3.4	Model 3; An iterative approach	28
3.5	Model 4: Toulmin and argumentation	29
3.6	Model 1 expanded: concepts of evidence and claims	30
3.7	Model 4: Toulmin and ‘looking forward and back’	31
3.8	Model 1 and ‘looking forward and looking back’	32
3.9	Model 1 and school practical work	32
3.10	Model 1 and scientific literacy	33
3.11	Model 5: Empirical work, argumentation and scientific literacy	37
4.1	An overview of the research	41
4.2	The ISEI variable	48
5.1	The ideas behind the teaching sequence	53

List of Tables

2.1	A section of the Concepts of Evidence associated with Design validity of investigations . .	12
2.2	Perceptions of the procedural component	14
2.3	How to teach, and what to teach	15
4.1	The sample	45
4.2	The age structure by year	45
4.3	School types attended	46
4.4	Number of A levels	46
4.5	Age and access route	46
4.6	Age and first generation university	47
4.7	The ISCO classification of occupations	47
4.8	The Goldthorpe categories	47
4.9	ISCO and Goldthorpe classifications	48
4.10	Age and Goldthorpe classification	48
4.11	SES and attainment at GCSE	49
4.12	SES and first generation university	49
4.13	Age and GCSE	49
4.14	The sub-samples which completed various of the assessments	50
5.1	The course: a brief overview	54
5.2	The course: Session 3	55
5.3	The course: session 5	56
A.1	The item weightings in the test score	63
A.2	Evidence pre-test: correlation matrix (N=137)	64
A.3	Evidence post-test: correlation matrix (N=137)	65
A.4	Evidence pre-test: item statistics (the Cronbach function in R)	65
A.5	Evidence pre-test: summary statistics (N=165)	65
A.6	Evidence pre-test: statistics for the post-test items (the Cronbach function in R)	65
A.7	Evidence post-test: summary statistics for the post-test items (N=165)	65
A.8	Evidence post-test: item statistics (N=137)	66
A.9	Evidence post-test: summary statistics (N=137)	66

Chapter 1

Introduction

1.1 The structure of the reporting of the research

There are a number of sources of information available to the reader:

1. a web site with links to what follows: <http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm>
2. this report, which details some of the background thinking at a level of detail not possible in a journal article
3. a series of journal articles which we refer to at various stages of this report and which are referenced on the web site
4. appendices to this report which include all the instruments we use, together with coding guidelines and reliability data where appropriate

1.2 The historical background

It is now some twenty years since the Assessment of Performance Unit in Science in England (see for instance Gott and Murphy [1987]) began to assess pupils' ability to carry out open-ended investigations. The tasks they used were very different from earlier varieties in the UK (such as those in Nuffield curricula) in that they gave the very barest of details and, importantly, were not heavily reliant on traditional substantive ideas (of force, or photosynthesis for example). Thus a task was presented in the style:

'Find out how the bounce of a squash ball depends on its temperature'.

It is clear from this that we are looking at the pupils' ability to generate reliable empirical evidence rather than an explanation based on elasticity, or gas pressure. This early work proved influential; for good and for ill many would argue. The National Curriculum for England which followed was heavily influenced by APU reports and, as a consequence, included open-ended tasks of this sort in its specification. We have specified the ideas we consider constitute the 'domain' of a procedural understanding: the ideas which are important for judging the validity and reliability of evidence in an investigation. (It is worth noting that 'procedural understanding' is a term used by other authors with different meanings. Some regard 'procedural understanding' as a term similar to meta-cognition; the awareness a pupil has of the procedures they are using and thinking about.) These procedural ideas, we have argued, are a necessary if insufficient condition for good experimentation. We have identified a long list of these ideas (which we have labeled 'concepts of evidence' for want of a better term) which constitute an element of a 'syllabus' for procedural understanding and are therefore of importance in a curriculum preparing pupils for future work in science.

More recently we have also argued that the same ideas are an important tool for the variety of scientific literacy which concerns itself with empowering people to challenge scientific data and the conclusions and decisions which follow from these.

These ideas have met with criticism from various quarters. Those criticisms that we know about can be summarised thus:

1. These ideas are not science, they are maths, or technology, or engineering. This, of course, may be true but gets us nowhere except into an arid dispute on subject boundaries.
2. The ideas are not important and may therefore have been ignored. We have attempted to answer this criticism by examining their role, or lack of such, in industry and in matters of public debate. Our conclusions are that they certainly are important; some more than others, Gott et al. [1999a], Tytler et al. [2001a], Tytler et al. [2001b], Ryder [2001]. We must stress here that they are an important component, not a panacea, and substantive ideas coupled with a driving interest in a particular issue are decisive.

And then we have a series of comments which are open to empirical research and to which we have addressed ourselves here.

1. The ideas are too hard for pupils (and even teachers it has been suggested).
2. The ideas cannot be taught effectively.
3. There is no evidence that understanding those ideas makes for better experimental work.
4. There is no evidence that such understanding gives people a better purchase on the sort of scientific data which is (supposed) to inform decisions on health, or environmental issues (for example).

On these last points we have, indeed, very little evidence. That evidence is very difficult to get. The dreaded league tables of schools which dominate educational ‘thinking’ in England put schools in a position where any innovation, which cannot be seen directly to contribute to better assessment results is taboo. We have been able to carry out minor exploratory work (Roberts and Gott [2006b], Gott and Roberts [2004]), but nothing on a scale which would allow for even the most rudimentary testing of most of the questions outlined above.

As a consequence, we embarked on a research programme which makes use of the fact that students at Durham University enrolled on a course for intending primary teachers can, and we argue should, be taught these ideas as part of their science modules. This allows us to construct the programme of research which is reported here. Of course, it is not ideal. The students are certainly in the upper half of the ability distribution; they are older, more committed and experienced. So transfer of findings to school pupils of 11-16/18 years of age is problematic. We also must recognise that the sample here is almost entirely made up of females, another reason not to attempt generalisations without very careful thought and much by way of caution.

Nevertheless, we have been able to carry out a reasonably comprehensive piece of research which covers the problems outlined above. We stress here, and in the theoretical chapter later in this report, that we do not attempt to accommodate our research to the, often conflicting, views present in the literature on practical work and, *a fortiori*, scientific literacy. The debate is so loosely focused as to allow any and every view, and as a consequence, goes nowhere. We set out our view in some detail in a subsequent chapter. The reader is then free to accept or reject it.

1.3 Methodological issues

In this research, we typically have sample sizes of around 50 to 100. This is too large for detailed qualitative work, although of course we use this approach on a subset of the students. But it is, arguably, too small for regression analysis, although there are many examples in the literature of such an analysis being used on smaller samples than this. However, we choose not to use conventional statistical tools here as our major analytical focus. Rather we choose to use the idea of Qualitative Comparative Analysis (QCA) championed by Ragin (see for instance Ragin [2007]) in the social sciences in the USA, and Cooper in education in the UK (Cooper [2005]).

This is an approach which is likely to be unfamiliar as yet to most readers and we have taken some care in associated papers to offer our explanation of its use and its costs and benefits.

Chapter 2

Background: the procedural component of a science curriculum

2.1 Introduction

‘At the core, science is fundamentally about establishing lines of evidence and using the evidence to develop and refine explanations using theories, models, hypotheses, measurements and observations’ (Duschl et al. [2006] chap 1 p 6)

Recent curriculum frameworks around the world have reflected science as more than a body of facts (see for example Duschl et al. [2006], Qualifications and Authority [undated], Curriculum Council [1998]). Students are now expected also to engage with the central role of evidence in science.

Open-ended investigations enable them to do this. Open-ended investigations are those in which students are unaware of any ‘correct’ answer, where there are many different routes to a valid solution and where different sources of uncertainty lead to variations in repeated data so that students reflect and modify their practice in the light of the evidence they have collected. The evidence produced, then, is messy rather than the laundered version common in practical work contrived to illustrate ideas to students. Such investigations play an important role in ‘inquiry’ based approaches to science curricula across the world (Abd-El-Khalick et al. [2004]). They reflect ‘science as practice’ and are important for understanding not just the practice of scientific investigation but also provide a context for understanding why science needs empirical evidence (Duschl et al. [2006]).

In this small scale research we seek to explore empirically the importance of students’ understanding on their ability to conduct open-ended investigations. The findings may help to inform curricula, specifically about what to teach, which have aims of developing students’ proficiency in investigations.

2.2 What understanding is important?

Our view is constructivist in that understanding requires the learner to construct meaning from ideas. In science the traditional curriculum emphasis and focus in classroom teaching has largely been on the content knowledge ideas of the subject matter such as mechanics in physics or genetics in biology, to name but two. In other words, we are dealing with the familiar factual knowledge, concepts, laws and theories of science. This is traditionally known as substantive or conceptual understanding.

The facts, concepts, laws and theories that contribute to this substantive understanding are, of course, themselves supported by empirical evidence or are subject to investigation. Science, therefore, encompasses more than just an understanding of the familiar factual knowledge, concepts, laws and theories. This traditional substantive understanding alone is not sufficient to describe the ideas of science. Another key curriculum component is concerned with the procedures of science. Yet this component is often down-played in curriculum documents, at least in terms of the number of pages devoted to its specification. The procedural component is often represented in curricula in terms of behavioural objectives (Duschl et al. [2006]) and is implicit in text books (Roberts and Gott [2000]); a performance emphasis rather than emphasising understanding per se.

The term ‘procedural knowledge’ is found in a number of places, not just in science, and implies ‘knowing how to proceed’; in effect, in science, a synthesis of manual skills, ideas about evidence, tacit understanding from doing and the substantive content knowledge ideas relevant to the context. In science education in the UK and in our work, the term procedural understanding has been used to describe the understanding of ideas about evidence, which underpin knowing how to proceed.

The term ‘procedural knowledge’ has been used to distinguish ideas about evidence from other more traditional substantive ideas. As a set of ideas, ideas about evidence should be thought of as further substantive or conceptual knowledge in science; we consider these ideas to be an important sub-set of the substantive ideas in science. However, in our use of terminology we have distinguished them from the more traditional substantive idea and have argued that a lack of these ideas prevents students exhibiting a procedural understanding.

This research questions whether understanding the traditional substantive content knowledge ideas of science is sufficient for students to be proficient in an open-ended science investigation. Or are other factors necessary or sufficient conditions? The specification and teaching of the traditional substantive ideas of science is not the focus of this paper. That is not to down-play its importance in science education. Our focus is on the procedural component of the curriculum. The procedural component of the curriculum has been conceptualised and articulated differently in the literature by researchers with different research agendas. We will briefly characterise these in Section 3 and position the rationale for this research. But first we need to delimit what we are discussing.

2.3 Other perspectives on the procedural component

Our focus is on the procedural component of the curriculum; how it can be perceived and the consequences of these different perceptions on how it is taught and assessed.

However, different research and pedagogical traditions are also concerned with the procedural component of science and we need to clarify what this paper focuses on with respect to others’ views.

2.3.1 Piagetian schema

We start with extensive research that has been strongly influenced by Piagetian psychology (Inhelder and Piaget [1958]); see for instance Toth et al. [2000], Klahr and Nigam [2004], Adey [1992], Kuhn et al. [1988], Schauble [1996]. The development of ‘higher order thinking skills’ are thought to be important in helping students better understand science. The development of strategies about control of variables and causal relationships, for example, are shown to be important for understanding science, particularly aspects of the procedural component. The focus of such work is on the development of appropriate schema in pupils so that they can become formal reasoners. Psychology-informed research has shown that students can be explicitly taught to develop this understanding; it is not just a ‘skill’ that develops only with practice (c.f. the ‘skills’ perspective developed in section 3.1). Klahr and Nigam [2004] demonstrated the effects of explicit teaching, which seemed to develop understanding which lasted much longer than in students who were just left to practice. Others have shown similar results (see Kuhn and Dean [2005], Shayer and Adey [1992], Chen and Klahr [1999] and Toth et al. [2000], for instance). However, in interventions from this psychological perspective, what is taught, such as the control of variables strategy, is seen not so much as part of science per se, thereby legitimising it within curriculum structures, but as an ‘additional’ psychological component necessary for developing understanding in science (Jones and Gott [1998]). In the UK, for instance, in the Piagetian-based CASE (cognitive acceleration through science education) programme’s interventions, what was taught was selected from the psychological perspective of developing formal reasoning. The interventions, which arguably addressed some of the procedural component of a science curriculum, were seen rather more as ‘brain training’ rather than the teaching of ideas about evidence that are integral to science. Jones and Gott [1998] contrasted this with the ‘understanding ideas about evidence’ perspective which underpins this research and is expanded in section 3.2: the crucial difference is that from the ‘understanding ideas about evidence’ perspective, the ideas of evidence are perceived to be important ideas of science, and as such can be selected to construct curricula, can be taught and can be assessed. Thus, although research from a Piagetian perspective has much to inform the procedural component in science education and there is overlap in what is taught,

the ‘understanding ideas about evidence’ perspective focuses on ideas integral to science and covers understanding of more ideas from science than does the psychology-focused development of schema.

2.3.2 Inquiry

The word ‘inquiry’ is prominent in literature about the procedural component, as is evident from an internet search on the term, but it is used in a myriad of ways. For instance, an inquiry (or inquiry task) can be similar to an open-ended investigation, as we have described above. Inquiry is also used as a term for a more psychology-focused process akin to ‘scientific thinking’. Inquiry-based curricula employ an inquiry-based teaching and learning approach. But even as a pedagogical approach the word can imply different things; from simple ‘discovery learning’ (often contrasted with direct instruction) to planned learning progressions which include more explicit teaching in the sequence of lessons (Duschl et al. [2006]). In the UK the inquiry approach was typified by Nuffield curricula (Jenkins [1979]) and emphasised learning substantive ideas through practical work and the US view is not dissimilar (see for instance Duschl et al. [2006]). Within nearly all the uses of the term inquiry is the view, implicit or explicit, that both the substantive and procedural components of the curriculum are inseparable; the two components are addressed together in inquiry with the resultant emphasis in practice being largely on the substantive ideas (see section 4.1.2). This is in contrast to the argument we develop in this paper where we have clearly distinguished the substantive and procedural components of the curriculum to structure the discussion. Inevitably there is an overlap between our work and research framed in terms of inquiry but the focus is different.

2.3.3 Nature of Science

Research and curriculum developments concerning epistemology and the nature of science (see for instance, Hart et al. [2000], Sandoval [2003], Osborne et al. [2003], QCA [2004]) address understanding of the way science works. The emphasis in work from this perspective is largely epistemological and sociological. We are not concerned with these emphases here, although there is a lot of common ground. The procedural component of the curriculum is important in the nature of science but much of the research comes from a more philosophical perspective. The concepts of evidence (see section 3.2), we would argue, are a sub-set of the ideas involved in such an understanding of the Nature of Science.

2.4 The Procedural component of the curriculum

The procedural component of the curriculum is concerned with ‘doing science’ (Hodson [1991]). Polanyi considered some forms of expert procedural knowledge, ‘know-how’, to be tacit and unable to be codified (Polanyi [1966]). From this perspective, expertise is more than the sum of the component parts. But what, at least, are these component parts? What do we need to teach students to help them to investigate? Two contrasting perspectives of how the procedural component is conceived can be characterised as ‘a skills’ perspective and ‘an understanding of ideas about evidence’ perspective. While we see these as two fundamentally different perspectives we recognise that some literature contains some elements from both. The perspectives taken will, in turn, influence both how the procedural component is taught and assessed. We will attempt to do illustrate this in the following account and it will be developed in a later chapter.

2.4.1 The skills perspective

The skills perspective is characterised by performance, often termed as ‘process skills’. The main characteristic of such a perspective is that the procedural component is thought to be learned by practice with repeated exposure: by doing lots of practical work that involves these skills it is assumed that students will improve their conceptions of how science works. The procedural component is largely implicit in teaching and any guidance given to students is simple exemplification of the process. An early version of this is typified by Science: A Process Approach developed from work by Gagne which identified isolated ‘process skills’ (American Association for the Advancement of Science 1967) which was followed by others

including, in the UK, Warwick Process science (Screen [1988]) which emphasised ‘process skills’ such as observing, classifying and interpreting. Research has shown that (Duschl et al. [2006] Chap8 pp2-3):

“children failed to develop meaningful understanding under science-as-process instructional programs . . . but its legacy persists in both policy and practice”

There is a danger, we would argue, that elements of that legacy can still be seen in curricula that either have procedural components specified as behavioural objectives, since these may be translated into classroom practice and assessment as just ‘doing’, or in curricula that emphasise using investigations as a pedagogical approach, a way of teaching, for mainly substantive understanding. In such pedagogical approaches the ‘doing’ of science is considered to be sufficient to meet the procedural component of the curriculum; students are considered to ‘discover’ the procedural element with practice.

2.4.2 An understanding of ideas about evidence perspective

An understanding of ideas about evidence perspective is a different way of conceptualising the procedural component of the curriculum. It is this perspective that underpins our research. The procedural component is seen to be underpinned by a set of ideas about evidence. It requires the learner to construct meaning, specifically about validity and reliability, from specific ideas about evidence. The focus is on a set of ideas that are an integral part of science and that can then be learned, understood and applied, rather than a set of skills that develop implicitly by practice. These ideas can be applied and synthesised in open-ended investigations, together, of course, with the traditional substantive ideas of science. (We also consider them to be important in empowered forms of scientific literacy, to enable students engage with scientists’ claims and scientific argumentation (Tytler et al. [2001a], Tytler et al. [2001b], Gott and Duggan [2007], Roberts and Gott [2007]) which we develop elsewhere in this report.) They are the ideas that underpin performance and therefore underpin curricula written with behavioural objectives.

‘Procedural knowledge (in the context of experimental work) will inform decisions, for example, when planning experimental investigations, processing data and using data to support conclusions’ (p1137 Buffer and Lubben [2001]).

We have referred to these ideas about evidence as ‘the thinking behind the doing’ (not to be confused with meta-cognitive notions of ‘thinking about one’s own thinking’) and have created a tentative list numbering some 80 or so of them which we have called the concepts of evidence¹. They serve, we argue, as a domain specification of ideas necessary for procedural understanding. From this perspective, the procedural component of a curriculum consists of ideas (in effect, a sub-set of substantive ideas) that form a knowledge-base of evidence that can be explicitly taught and understood, in a similar way to the more traditional substantive elements in the curriculum. The concepts of evidence cover some 19 areas ranging from fundamental ideas of causation and association, through experimental design, data analysis and interpretation to validity and reliability as overarching touchstones of evidence ‘quality’.

Table 2.1 is illustrative of part of the section of the Concepts of Evidence associated with the validity of investigation design. (<http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm>) and are described in more detail in a later chapter.

¹<http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm>

Table 2.1: A section of the Concepts of Evidence associated with Design validity of investigations

Topic	Understandign that:	Example
Fair test	A fair test is one in which only the independent variable has been allowed to affect the dependent variable.	A laboratory experiment about the effect of temperature on dissolving time, where only the temperature is changed. Everything else is kept exactly the same.
Control variables in the laboratory	Other variables can affect the results of an investigation unless their effects are controlled by keeping them constant.	In the above experiment, the mass of the chemical, the volume of liquid, the stirring technique and the room temperature are some of the variables that should be controlled.
Control variables in field studies	Some variables cannot be kept constant and all that can be done is to make sure that they change in the same way.	In a field study on the effect of different fertilisers on germination, the weather conditions are not held constant but each experimental plot is subjected to the same weather conditions.
Control variables in surveys	The potential affect on validity of uncontrolled variables can be reduced by selecting data from conditions that are similar with respect to other variables.	In a field study to determine whether light intensity affects the colour of dog's mercury leaves, other variables are recorded, such as soil nutrients, pH and water content. Correlations are then sought by selecting plants growing where the value of these variables is similar.
Control group experiments	Control groups are used to ensure that any effects observed are due to the independent variable(s) and not some other unidentified variable.	In a drug trial, patients with the same illness are divided into an experimental group who are given the drug and a control group who are given a placebo or no drug.

The concepts of evidence attempt to specify the procedural knowledge base underpinning the collection and evaluation of evidence; the things students need to know to be able to judge validity and reliability, either from their own investigation or that reported by others. The ideas are a toolkit of ideas integral to the planning and carrying out of practical investigations with understanding (rather than as a routinised procedure, Roberts and Gott [2003]). We would argue that they are necessary but not sufficient for an investigation: manual skills, the tacit knowledge described by Polanyi (Polanyi [1966]) and, of course, the more traditional substantive ideas of science are important as well.

Essentially an investigator needs to be sure of the reliability and validity in each 'layer' when they make and defend a claim from empirical evidence. We have argued the same ideas are used by others when evaluating others' claims, whether that is in formal scientific peer review, in public engagement in science or in scientific argumentation in the classroom (Gott and Duggan [2007], Roberts and Gott [2007]). Since data is at the heart of science investigations we start at the centre of the diagram. For example:

- for each datum we need to consider the quality of any observation or reading taken. If the datum is invalid or unreliable, the whole investigation is invalid;
- for a data set we need to consider whether sufficient repeated readings have been taken to capture the variation and enable us to trust the data and whether the data has been validly summarised;
- when seeking relationships between variables the validity of the design must be considered as well as the interpretation of the data taking into account any uncertainty arising from the data collection;
- comparison with other sources that may have influenced the design or conduct of the investigation requires judgment on the validity and reliability of others' work, and

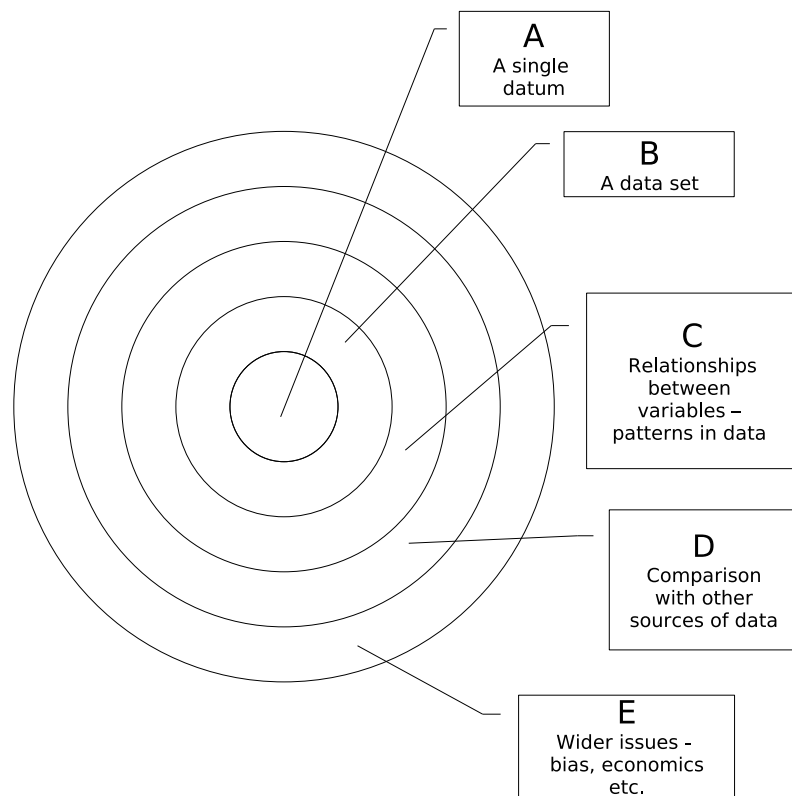


Figure 2.1: Bullseye

- the reliability and validity of the whole investigation must also take account of wider societal issues that might have influenced any stage of the work.

The concepts of evidence include ideas about the uncertainty of data (as taught and researched by Buffler and Lubben [2001]). They also include ideas important to understanding measurement and data processing, presentation and analysis which may be considered to be part of a mathematics curriculum but which are essential for understanding evidence. In section 2.1 we have pointed to overlaps with other research perspectives.

2.5 Teaching the procedural component

The way in which the procedural component of the curriculum is conceived affects decisions about how it might be taught and assessed. Assessment practices often have a backwash affect on teaching. We will start by considering teaching without such negative influences. Assessment will be discussed in a later section and there we will also address how it might distort teaching.

Approaches to teaching the procedural component are not only influenced by whether it is seen from a skills or an understanding about evidence perspective but are also affected by pedagogical decisions about whether substantive and procedural components should be taught and learned together or separately. We illustrate this in table 2.2 and then discuss each cell in turn. Again, these distinctions are somewhat blurred in practice. The purpose of such characterisation is to position our teaching for this research in relation to other approaches in the literature, table 2.2.

Table 2.2: Perceptions of the procedural component

	Substantive and Procedural components taught separately	Substantive and Procedural components taught together
	Separately	Together
Skills	1	2
Understanding of ideas about evidence	3	4

2.5.1 Teaching from the skills perspective (cells 1 and 2)

From the skills perspective ‘practice makes perfect’, so teaching involves repeated exposure to practical work to the exclusion of other methods or explicit teaching of the underpinning ideas.

2.5.2 Procedural component separately, (cell 1)

Approaches like this, such as SAPA (American Association for the Advancement of Science 1967)

‘treated . . . science as a series of disembodied ‘skills’, [which] were often carried out as meaningless procedures (Baroody et al. [2004], Mintzes et al. [1997]) ... Ultimately ...children failed to develop meaningful understanding under science-as-process instructional programs, and researchers recognised how little these domain-general ‘skills’ actually generalised. Another criticism was that they were based on faulty developmental assumptions about children’s reasoning and learning capacities (e.g., that young children are concrete rather than abstract thinkers and capable only of observation rather than explanation’ (Duschl et al. [2006] Ch 8, p2).

2.5.3 Procedural and substantive components together, (cell 2)

The skills perspective, we argue, is the default position taken by teachers and text developers (Roberts and Gott [2000]) if explicit instruction about the procedural component is not provided when both the substantive and procedural ideas are taught together. For instance, when ‘processes’ are identified and students conduct practical work to illustrate substantive ideas, or to support their ‘discovery’, the procedural component can be down-played by the careful guidance of the teacher so that the experiments ‘work’ (in terms of demonstrating the anticipated substantive ideas) and the students may not have to worry about design decisions or interpreting messy data. For instance, the ideas associated with the quality of a measurement which are central to any investigation, such as instrument calibration, precision, resolution and the resultant systematic and random uncertainty, may not be even encountered as ‘skills’ in an investigation where the student is given no choice of instrumentation and where their attention is focused on the substantive ideas of science (Sandoval [2003]; Sandoval and Reiser [2004]) Many pupils left to ‘discover’ the procedural component do not develop better understanding (Klahr and Nigam [2004]). There is some evidence that this implicit ‘skills’ approach works for some students. It has been shown that extended practice in a context over several weeks was successful (Kuhn and Phelps [1982]) and that frequent engagement with investigations lead to developments (eg Kuhn et al. [1992]). Key to this is the opportunity required for such practice, which in a classroom context will need long periods of time.

2.5.4 Understanding of evidence perspective (cells 3 and 4)

Once we recognise that there is a set of ideas underpinning evidence, then decisions about how to teach these ideas can be made on a different basis from the repeated practice associated with the process skills perspective.

The discussion moves away from teaching performance and moves onto ways of teaching the ideas about evidence, with practical work as just one of many ways in which they could be taught (just as both practical and non-practical work can be used to teach the substantive ideas of science, shown in Table 2.3). Such an approach may well be less demanding of teaching time and therefore more efficient.

Table 2.3: How to teach, and what to teach

	Practical activities which might include:	Other non-practical activities which might include
Substantive ideas	Observations of objects or events and their classification, illustrative practicals, ‘discovery’ learning and enquiry practicals	Didactic teaching, active learning using texts, discussion, presentations, the use of models
Procedural ideas	Whole or parts of investigations in field and lab, illustrative practicals, basic skills practicals	Didactic teaching, active learning using texts, discussion, presentations, the use of second hand data, evaluating investigations

Various researchers have shown that explicit teaching of the ideas which contribute to a procedural understanding does develop understanding and it seems to be effective. For example, Buffler and Lubben [2001] successfully taught students about ‘uncertainty’. Klahr and Nigam [2004] demonstrated the effects of explicit teaching, which seemed to develop understanding which lasted much longer than in students who were just left to practice. Some who explored the task on their own developed understanding, but far more did so with explicit instruction. Kuhn and Dean [2005] also found improvement with practice but greater improvement when students received instruction. Chen and Klahr [1999] compared direct explicit teaching of control variables (whereby students were specifically taught about ‘control of variables strategies’ using didactic teaching, examples and questioning), with ‘implicit training’ (which consisted of using probes during student experimentation) and a control group (who had no teaching at all). Direct teaching had a huge impact (34% use of CV strategy before teaching, 61-65% after even 7 months) and implicit and control groups made no gains. Similar effects in classroom research eg Toth et al. [2000] increased the number of students who used the ‘control of variables strategy’ when evaluating others’ research (up from 28% to 76%).

2.5.5 Procedural component separately, (cell 3)

This is the approach we have used in the intervention in this research. It is described in more detail later in this report and in Roberts and Gott [2007]. We describe it in some detail to illustrate Cell 3.

Students’ understanding of ideas about evidence is developed by explicitly teaching them the concepts of evidence required to conduct an investigation. Ideas can be introduced in a sequence to progressively introduce students to the ideas they might require. Since the teaching aims to develop these ideas about evidence, and does not include any substantive subject knowledge, the examples used for teaching, practical work and the context for the open investigation are deliberately very simple so that the substantive ideas required to solve problems are kept to a minimum; they are ‘knowledge-lean’ tasks in Kuhn’s terminology (Kuhn and Phelps [1982]).

The ideas about evidence are the focus and ‘learning outcomes’ of the teaching using both practical and non-practical work. The range of teaching activities include:

- The use of text (Gott and Duggan [2003a]) to introduce and explain ideas,
- Exploratory practicals to familiarise students with the ideas about evidence,
- Practicals to illustrate specific ideas about evidence,
- Parts of investigations where students can apply some of the ideas they have learned,
- A whole guided investigation,
- Whole class and small group discussions focused on the ideas about evidence,
- Video,
- Structured worksheet activities mainly in the form of question sheets.

2.5.6 Procedural and substantive components together, (cell 4)

The underpinning rationale behind this cell is summarised by Duschl et al. [2006] Chapt 2, p9: "

‘The strands are not independent or separable in the practice of science, nor in the teaching and learning of science’

. While the practices of scientists synthesise both substantive and procedural components, table 2.2 shows that such a synthesis is not inevitable in the teaching of the components. However, the educational rationale for this cell may be influenced by calls for teaching to be based on ‘meaningful questions’ which may better reflect ‘authentic’ science and may build on students’ own observations and interests.

But there is a fine line between this idealised situation and the practices typified by Cell 2. If the ideas required for a procedural understanding are to be taught so that students can develop understanding there must be a careful sequencing of ideas so pupils are capable of investigation (Sandoval and Reiser [2004]). There may be tensions between competing demands for separate progressions in ideas for substantive understanding and the ideas about evidence. In our experience, the predominant ideas in the curriculum, the substantive ideas, become the focus of planning and teaching.

2.6 Assessing the procedural component

In this section we will discuss summative assessment, concentrating on the distinctions inherent in the ‘skills’ and ‘understanding of ideas about evidence’ perspectives of the procedural component.

2.6.1 Assessment from the skills perspective

Assessment here focuses on performance in whole investigations and tasks that focus on particular ‘process skills’.

Paralleling the developments to teach process skills were assessment schemes such as The Assessment of Practical Science (Bryce et al. [1983]) and the Graded Assessment in Science Project (Davis [1989]) which were both based on the assessment of ‘process’ in the context of isolated practical skills.

Woolnough [1991] pointed to the problems associated with assessing the quality of an investigation when:

‘quality . . . is recognisable but indefinable’ (p8).

He warned of the trivial exercises that would follow an insistence on tight reliability against a closely prescribed set of criteria and suggested that

‘when we attempt to describe a series of steps along which students should progress when tackling an open-ended scientific investigation, we run into fundamental problems concerning the very nature of scientific activity - it does not progress along such tidy lines’ (p9).

Research has shown that there are significant problems with the reliability of the assessment of performance of investigations. Students’ performance is affected by a number of factors (summarised by Duggan et al. [1992]) including the subject matter of the investigation, the setting or context (lab or field, for instance) and the ‘procedural complexity’ such as the variable structure and type of variables, the degree of interaction with the apparatus required and the openness of the task. Solano-Flores et al. [1999] attempted to reduce the variation due to subject matter and context in assessment tasks by creating writing ‘shells’ that took account of this ‘procedural complexity’ in the construction of assessment tasks but found that there was still considerable variation in pupils’ performance on tasks of apparently similar demand. So clearly there are a number of problems with performance assessment.

We have argued elsewhere (Duggan and Gott [2002], Roberts and Gott [2003], Duggan and Gott [2000]) that practical assessment by performance in investigations presents intractable problems in a mass assessment system catering for all abilities such as GCSE (the examination at the end of compulsory schooling at 16 in England and Wales). The evidence points to the need for as many as 5 to 10 open-ended tasks set in different contexts before we can be confident of the reliability of the resulting ‘score’ obtained from the reporting of such work (Gott and Duggan [2002]). Establishing an appropriate reliability in, say, 2 or 3 tasks, which is all there is time for, requires that they be ‘routinised’

(Roberts and Gott [2003]) whereby performance can be reliably assessed only if everyone performs in the same way, or converted into assessments of substantive ideas - which are already assessed in traditional exams - by giving credit to the predictions, hypotheses and interpretations of theory (Duggan and Gott [2002]). This has been shown to have had a negative backwash on teaching in the UK. The House of Commons Science and Technology Committee (The House of Commons and Committee [2002]) made a number of trenchant criticisms of the assessment arrangements for GCSE. They castigated coursework, which consists of whole and parts of investigations, as ‘boring and pointless’ (p50) and practical work as being something that ‘students see little point in carrying out . . . where they already know the result and are just expected to follow instructions to reach that end.’ (p20) They went on to say that ‘The way in which coursework is assessed for GCSE science has little educational value and has turned practical work into a tedious and dull activity for both students and teachers.’ (p21)

The validity of the assessments, as reflecting the diversity of scientific practice, is then called into question (Donnelly et al. [1994], Donnelly [2000], Duggan and Gott [2002], Watson et al. [1999]).

2.6.2 Assessment from an ideas about evidence perspective

From this perspective, assessment is about the ideas about evidence that students know, understand and apply. These ideas can be assessed in a variety of ways, just as they can be taught using different approaches. Two such forms of assessment are described below, both of which were used in this research.

Relatively short efficient written tests and probes can be used that target the concepts of evidence thus assessing the understanding of the underlying procedural ideas. Such written probes are used extensively in research (for example Gott and Roberts [2004]; Germann et al. [1996], Germann and Aram [1996]; Buffer and Lubben [2001]; Gotwals and Songer [2006]). They have only recently been introduced into mass assessment systems in the UK (Roberts and Gott [2006a]). The evidence test employed in this research was a written test of students’ understanding of the ideas about evidence.

The application of the concepts of evidence, as they are transferred to inform decisions used in an open-ended investigation, can also be assessed. Assessment of an open-ended investigation is not then viewed as being about the performance, per se, but about the ideas about evidence that have been applied to the problem. An advantage of assessing the ideas of evidence is that they underpin all types of investigation and thus reliable assessment is not dependent on everyone performing in a similar way (Roberts and Gott [2003]). This is the approach we took to assessing the open-ended investigations in this research.

We assessed the ideas in students’ written accounts of the investigation. This can be considered a suitable surrogate for direct observation (Baxter et al. [1992], Gott and Murphy [1987], Welford et al. [1985]) which would have been too time-consuming. While Haigh [1999] found that students’ written accounts did not necessarily reflect the subtleties of students’ ideas we encouraged students to write about ‘the thinking behind their doing’ rather than write a more formal ‘apparatus, method, results conclusion’ account. Martin [1993] argues that a research report is an opportunity to persuade the reader of the reliability of the scientific claim, and we explicitly encouraged the students to justify their decisions and reasoning in a coherent account. This approach to writing has been adopted successfully by Toh and Woolnough [1994]. In the marking of the written accounts of the investigations, credit was given to students’ accounts where the concepts of evidence were explicitly and correctly applied. In our research which focused on understanding of the ideas about evidence, no credit was given to substantive understanding as such, only where it informed the procedure, i.e. in the selection and operationalisation of the variables etc. That is, we gave no credit to explanations of mass and gravity or forces and friction which were not the focus of the task.

2.7 How these ideas inform our research

Having outlined our position with regard to different perspectives in the literature it is time to consider how they inform the research reported in this paper. Summarising extensive literature on investigations, Duschl et al (Duschl et al. [2006], Chapter 5, pp1-2) suggest that:

‘Scientific investigation, broadly defined, includes numerous procedural and conceptual activities, such as asking questions, hypothesizing, designing experiments, making predictions,

using apparatus, observing, measuring, being concerned with accuracy, precision, and error, recording and interpreting data, consulting data records, evaluating evidence, verification, reacting to contradictions or anomalous data, presenting and assessing arguments, constructing explanations (to oneself and others), constructing various representations of the data (graphs, maps, 3D models), coordinating theory and evidence, performing statistical calculations, making inferences, and formulating and revising theories or models.’

Curriculum documentation that typifies the procedural component from a skills perspective, which we would argue is the perspective adopted by most curricula, seems to assume that students can successfully conduct an investigation if they are taught just substantive ideas and are given opportunities to practice investigations. It is easy to see that substantive ideas are important in investigations that have a substantive context or demand (‘knowledge-rich’, Schauble [1996]). But is substantive understanding enough? If an investigation has low substantive demand (it is ‘knowledge-lean’) what understanding is important? Is substantive knowledge sufficient?

In part of our research, we seek to explore the importance of students’ substantive understanding and their understanding of ideas about evidence on their ability to conduct open-ended investigations. The evidence with respect to the importance of these two types of understanding in conducting an investigation is not clear cut. Previous research (Gott and Murphy [1987]; Erickson et al., 1992; Millar et al. [1995]; Ryder and Leach [1999]) points to both playing a role, but these findings don’t give any insight into whether either or both are necessary or sufficient conditions for being able to carry out an investigation.

Research from the psychological perspective into the role of prior substantive knowledge and beliefs about causal mechanisms has been shown to affect hypothesis formation, the design of investigations and in the evaluation of evidence (summarised in Duschl et al. [2006] Chapter 5). Schauble [1996] found that, from a developmental point of view, children’s ability to investigate in ‘knowledge-rich’ contexts developed in line with both their understanding about the procedures of science and their substantive knowledge. These changes appear to bootstrap each other, so that appropriate knowledge supports the selection of appropriate experimentation strategies, and systematic and valid experimentation strategies support the development of more accurate and complete knowledge. Gotwals and Songer [2006] used a multi-dimensional modelling based on Rasch that identified 3 dimensions in students’ ‘inquiry’: content (substantive) knowledge, creating scientific explanations (which includes aspects such as identifying claims and evidence for them and associated reasoning), and interpreting data (which includes reading a table and graph and drawing conclusions from it). It is also pertinent to note that this research, along with most other work in this area, is based on post hoc surveys rather than lengthy interventions of the kind we describe here. Within the English school curriculum, it is not straightforward to disentangle the role of substantive understanding and understanding of ideas about evidence since there is little time devoted to the pupils’ conducting their own experiments (The House of Commons and Committee [2002], Roberts [2004]).

We are using a sample of undergraduate students in the UK who carried out such an open-ended investigation and wrote up their procedure and findings. We do not make any claims as to the generalisability of our findings, rather, we present this as a case study which may serve to give an indication of the respective roles of substantive understanding and an understanding of the ideas about evidence in these particular investigations and also to illustrate a possible approach to the problem of identifying necessary and sufficient conditions associated with successfully conducting an investigation. We are aware that context effects (Solano-Flores et al. [1999]) will play a part in investigations, in that the students’ performance will depend on whether the investigation they are asked to carry out is linked to the substantive subject matter they have recently been taught or to a particular interest they may have. We are also aware that students’ prior beliefs, and individual differences in the strength of those beliefs, about the tasks may influence observations the students make, the claims students make and their search for evidence (Chinn and Malhotra [2002], Kunda [1990]).

In this research, we propose (below) a simple typology of investigations that seem to require different degrees of substantive understanding for their successful solution (knowledge-lean through to knowledge-rich). In our previous research we have taken the view that understanding about evidence is a necessary ingredient in any open-ended investigation, regardless of the kind and amount of substantive understanding required (as well in scientific literacy more generally, but that is not the focus of this section),

but is it? And if it is necessary, is it sufficient, in an investigation where there is not a high degree of substantive understanding required?

Two of our research questions, with respect to our sample of undergraduate students, are:

- What are the necessary and sufficient conditions for success in an open-ended investigation?
- Are the conditions the same in two investigations that involve different degrees of substantive knowledge?

2.8 A simplified typology of tasks

To start with we introduce a simple typology of open-ended tasks within which to contextualise the research. Of course the boundaries between types are artificial to some degree. Gotwals and Songer [2006] created tasks with 2 dimensions: the difficulty of the substantive (content) knowledge and the difficulty of the procedural component. Our typology categorises tasks along the substantive (content) knowledge dimension. The open-ended tasks are similar to ‘hands-on’ tasks as described in Solano-Flores and Shavelson [1997]. They identified tasks with ‘high and low inquiry level’: our tasks posed a problem with just one independent variable - a ‘low inquiry level’ (Types 1 and 2 are like Solano-Flores and Shavelson [1997] ‘comparative investigation’ with a categoric independent variable; Type 3 had an independent variable with continuous values). The tasks asked students to solve the problem (another aspect of ‘low inquiry level’) but provided no guidance (an aspect of ‘high inquiry level’) and by providing students with all the available equipment in the lab from which they had to choose demanded that they make all their own decisions about design and measurement (‘high inquiry level’).

2.8.1 Type 1: Low on substantive understanding (Knowledge-lean)

Here we are talking about tasks which, while being recognisably scientific, are not heavily dependent on substantive theoretical structures. The example we use in this paper will serve to illustrate our point. The task is to determine:

‘Does the material of a ‘helicopter’ affect the time it takes to fall?’

A ‘helicopter’, made from paper or card, falls in much the same way as an ash or sycamore seed. We see that, to carry out this task, the student will require very little in the way of substantive understanding; on the face of it at least. They will rely on simple measurements of time and distance. We must stress here that explanation of the results does depend very much on theory but, we argue, a lack of such theory will not preclude a sensible empirical investigation.

2.8.2 Type 2: Substantive understanding plays a helpful part

An example of this type might be:

‘How much do different surfaces affect how easy it is for a shoe to slide?’

Whilst it is possible to carry out this task with little or no understanding of force and friction, our experience over a number of years, as well as common sense, tells us that such an understanding will lead to a better focused and more efficient solution. More details can be found at Glaesser et al. [2009].

2.8.3 Type 3: Substantive understanding is required

A ‘knowledge-rich’ task such as:

“How does the force experienced in a collision (of a car) depend on the design or size of a crumple zone?”

Without sophisticated accelerometers, investigating this task will require the use of Newton's laws, the equations of motion and ideas of potential and kinetic energy. A reasonably elegant solution, for instance, might be to allow a mass (the wall that the 'car' will hit) to fall from a known height onto a model crumple zone. Here application of ideas of energy can be used to show that the average force in the collision is the ratio of drop length of the mass to the crumple deformation multiplied by the weight (in N) of the car. After that, the task can be defined and the experimental phase takes over.

In all three tasks, however, there is reliance on procedural understanding to decide on the values to be used for the various independent variables, the number of repeats required, how to handle the repeat data and so on. The role of substantive understanding in investigations is explained further in a later section.

Chapter 3

The theoretical framework

3.1 Introduction

As we noted in the introductory chapter, our research interest lies in the framing of ideas related to evidence in science and how they might be taught and assessed, used to solve practical problems and act as a structure to look at issues to do with evidence in scientific literacy.

The debate on practical work in science, and more recently on scientific literacy, has not been notable for its clarity. The area is so broad and diffuse that almost any topic in science education can be included in one or the other of these debates. Our approach is, therefore, somewhat different. Rather than attempt to identify ideas from within the literature, which we have tried to do but with no success, we choose to define a framework from somewhere a little nearer to first principles. This has the advantage, we hope, of internal coherence and clarity, but the disadvantage (if such it is) of failing to deal with all possible points of view.

Our framework will be defined and illustrated in what follows. It is basically built up from a series of descriptive models which serve to frame the ideas we consider to be important and to act as a guide to research design and data analysis.

3.2 Models as descriptions and tools for analysis

In this section we shall introduce and develop, where necessary, a series of models. Their purpose is to reduce complex issues to something that is a little easier to grasp. These are essentially reductive and descriptive in nature. Their secondary purpose is to provide a focus for analysing the large volumes of data we have collected. They come in distinct sections. We shall deal with the underlying ideas behind evidence in science and its place as a stand-alone topic of some importance, its role in the accumulation and defence of evidence from practical investigative work in science, and as the basis for key elements of scientific literacy related to the development and analysis of an argument based on that evidence. We deal with each of these in turn below.

3.3 Defining terms

Before launching into the framework, we shall define a number of terms as we intend them to be understood in what follows.

Skills: We take there to be two distinct types of ‘skills’. Those associated with the manipulation of apparatus - the correct use and reading of a thermometer, for instance, or the correct assembly of more complex pieces of apparatus such as a potometer - and those ‘paper and pencil’ skills which were once ideas (constructing a table or a graph) but which have become second nature through repetition. Clearly there is no clear line between graphical and tabulating skills and concepts (both of evidence and substantive ones such as knowing the units for force) but the term is helpful and much used and so we shall follow suit.

Substantive understanding: Here we mean an understanding of the ideas in science that are well understood and comprise the underpinning structure of what is known. So we include Newton's laws, momentum, photosynthesis, substance and so on.

Procedural understanding: Here we mean an understanding of the set of ideas that underpin the collection, validation and analysis of evidence. We do not intend, thereby to suggest that substantive ideas are *not* so used of course.

Concepts: We take these to be the ideas that are necessary for someone to be said to have a good substantive understanding.

Concepts of evidence: These are the ideas of reliability, validity, error, uncertainty etc. which are necessary for a good procedural understanding. We have coined the phrase to distinguish these ideas from those in 'concepts' above, but in fact, of course, they are no different in kind, only in application.

3.4 The Concepts of Evidence: model 1

The most fundamental theoretical structure underpinning this research is what we have called the 'concepts of evidence'. We have advanced the idea that the 'what to understand' in science (the ideas used by scientists that can be considered part of a curriculum) can be seen to have two major components. The first is the familiar one of the substantive concepts underlying each of the science disciplines and, by way of illustration, covers such ideas as force, substance and photosynthesis. The second concerns the procedural ideas associated with the collection and verification of scientific evidence. Here we have two overarching ideas - reliability and validity. Sitting underneath these are a series of distinct ideas to do with experimental design, measurement, data representation and interpretation, and the appropriate statistical tools. We have generated an extensive list of these ideas, (together with a list of enabling skills, which is not the focus of this research). The concepts of evidence list has been quite extensively validated in university research, science-based industry and in public engagement with science (Tytler et al. [2001a], Tytler et al. [2001b], Roberts and Gott [1999]).

The concepts of evidence are a tentative list of the ideas that underpin procedural understanding. We do not suggest that every science context draws on all of them; but central to all evidence in science are the two overarching ideas of validity and reliability. The concepts of evidence can be found as an appendix to this document. There is no space to reproduce it here but it can be found as a pdf file in folder which accompanies this report or at our web site¹. We suggest that it can be used in a variety of ways:

1. As a basis for detailed research or curriculum development, in which eventuality the extensive detail and examples are a necessary element to be considered
2. As an aid to understanding our framework and research, in which case the reader might find it useful to have one or two pages from the list available to refer to whilst reading this document.

This list turns out to be very useful when considering the science curriculum; the specification of these ideas provides us not only with a domain specification for both teaching and assessment, but also a structure for this research. Some may view the detailed list of ideas as implying a very reductionist approach. We do not intend this. By listing the ideas, which together can be constructed into a procedural understanding, we 'map out' the knowledge-base, reducing any possible ambiguity inherent in more general descriptions. The concepts of evidence are *ideas* and as such can be considered in a similar way to the substantive ideas in science with which we are more familiar: we can consider, for instance, how they are taught, how well they are understood and applied in novel situations, the characteristics of student success and how we might assess them.

In an investigation to solve a problem for which a student doesn't know the answer, the student would need to make many decisions:

¹<http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm>

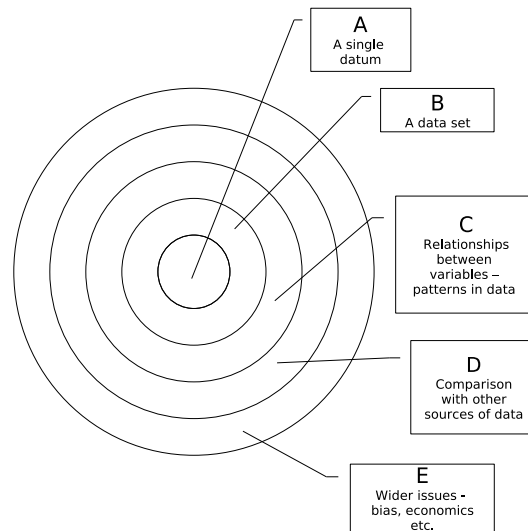


Figure 3.1: Model 1: the bullseye

- Exactly what is the question we are going to investigate and how do we design a valid investigation to answer it?
- What do we need to measure to give us data, the reliability of which must be open to judgment?
- How will we take the measurements, and how many will be needed?
- What is the most valid way to analyse the data?
- What conclusions can be drawn, taking into account the reliability and validity of the investigation as a whole?

The concepts of evidence are the ideas used to make such decisions in an investigation. Figure 2.1 summarises these as a nested set of ideas. The extensive list of the Concepts of Evidence is unwieldy for some analysis purposes (but critical for curriculum definition purposes). We have grouped them together into subsets, or layers, represented in the following bullseye diagram figure 3.1.

The concepts of evidence are the ideas used to make decisions in an investigation - the ‘thinking behind the doing’. Essentially we need to be sure of the reliability and validity in each layer. For example:

- for each datum we need to consider the quality of any reading taken;
- for a data set we need to consider whether sufficient repeated readings have been taken to capture the variation and enable us to trust the data;
- when seeking relationships between variables the validity of the design must be considered as well as the interpretation of the data;
- comparison with other sources that have informed the research requires judgement on the validity and reliability of others’ work and
- the reliability and validity of work must also take account of wider societal issues that may have affected the quality of the evidence.

These ideas should be seen as free standing and as elements in a science curriculum framework. They are useful in their own right, but we would argue they really come into their own when carrying out a practical investigation or when examining claims, in the press or elsewhere, about a science-related issue.

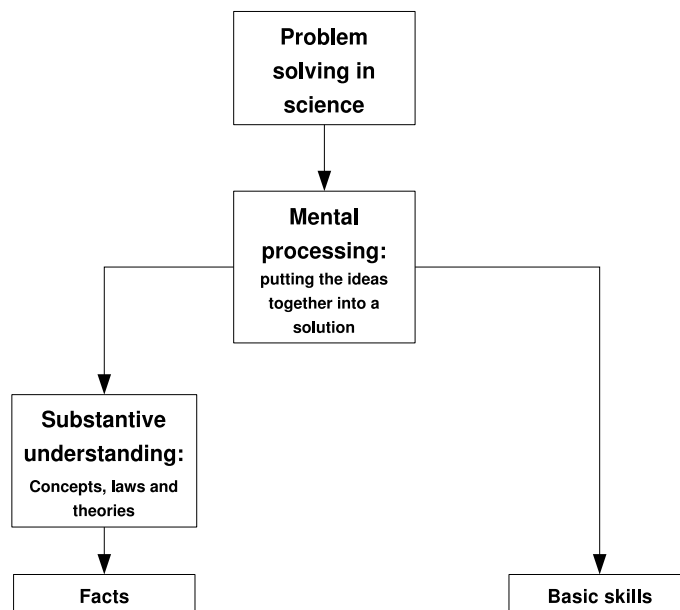


Figure 3.2: A skills approach within an explanatory focus

3.5 Evidence in investigations

We have articulated the concepts of evidence as the knowledge-base of a procedural understanding. These are the ideas which, together with the more familiar substantive ideas of science, are required to solve problems in science. Correctly applied, along with substantive ideas of course and of which more later, they result in data and a conclusion - a claim. We have indicated that the concepts of evidence are the ideas required to make decisions in an investigation. The literature in this area very often refers to practical work in science as involving a rather vaguely defined set of ‘skills’ or ‘processes’. We need now to distinguish these approaches from the way we are thinking about the investigations.

3.5.1 A skills approach to investigative work

To begin with we turn to a model we have used extensively to focus our thinking. In the UK national curriculum, a view has developed of investigations as a set of skills which, implicitly, can be honed by practice alone. This ‘skills’ approach is summarised in Figure 3.2 and is in contrast to our ‘ideas’ approach summarised in Figure 3.3. The position taken by the ‘skills’ approach is typified by a performance model. Pupils practice and perform ‘skills’, which would include not only handling apparatus and organisation skills but also ‘higher order investigative skills’ such as planning, measurement, observation, data presentation and so on, in the context of ‘action’ in a practical context. At the extreme, the assumption is that practice increases familiarity with these skills and that no specific understanding is required. It is the doing of science that matters and, what is more, the doing is there primarily to help in the understanding of substantive ideas and explanations stemming from them. The purpose of the investigation is either simply to carry them out, that being regarded as a ‘good thing’, to develop other ‘skills’ such as team work, or to establish or reinforce the substantive ideas which sit underneath the task.

3.5.2 An ‘understanding’ approach

We take the view summarised in Figure 3.3. This model (model 2) suggests that effective problem solving, the results of which may support and extend our understanding of a substantive idea, involves an interaction of both conceptual and procedural understanding. By procedural understanding we mean specifically ‘the thinking behind the doing’ of science which we regard as an understanding of the knowledge-base of the concepts of evidence required for deciding how many measurements to take, over what range and

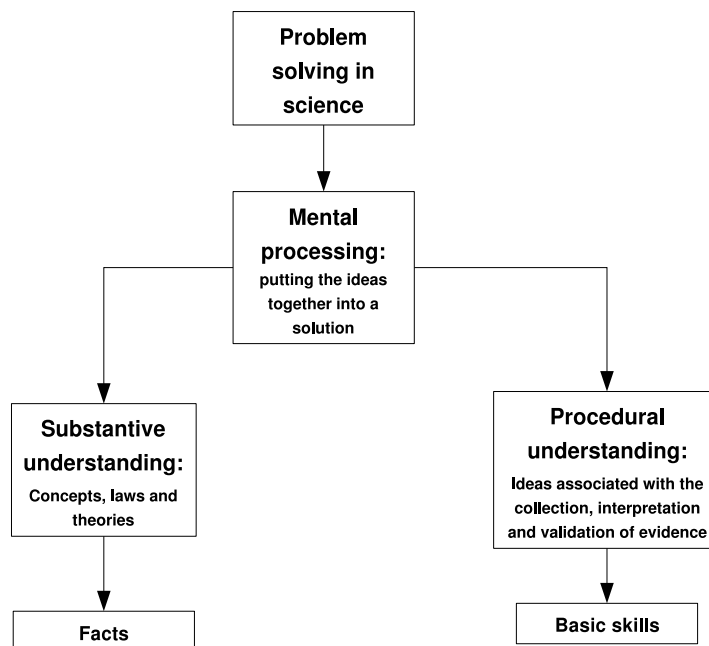


Figure 3.3: Model 2: An ‘ideas’ approach within a problem-solving focus

with what sample, how to interpret the pattern in the resulting data and how to evaluate the whole task. The underlying ideas have been discussed above as the ‘concepts of evidence’. But it also involves putting those ideas together into a set of tactics which deliver the strategic objective of solving the problem. The Concepts of Evidence, then, should be seen as necessary but not sufficient.

It is worth noting that ‘procedural understanding’ is a term used by different authors with different meanings. Some regard ‘procedural understanding’ as a term similar to meta-cognition; the awareness a pupil has of the procedures they are using and thinking about, rather than the understanding constructed from the specific knowledge-base of the concepts of evidence.

3.5.3 The role of substantive ideas

This is perhaps best illustrated in three examples.

1. **‘Helicopter’:** This investigation requires the solver to make a simple ‘helicopter’ - rather like an ash key - which then spins to the ground. The question asks them to investigate how the material it is made from affects the time for it to fall.
2. **‘Shoe’:** Here the students are asked to investigate how different surfaces affect how easy it is for a shoe to slide.
3. **‘Crumple zone’:** In this case (not used in our research project here by the way) the investigation focuses on how the forces experienced in a collision of a ‘car’ depends on the design of the crumple zone.

In the second, substantive ideas play a critical, but not absolutely necessary part. In the first they do not. In the third they are essential. But first we amplify a little what we are going to intend by the term ‘substantive concepts in science’. It is obvious what we are to *include* here; force, substance, photosynthesis and so on. But it is useful to define an idea by examining its negation. What, then, is to be excluded. We take the view that, for instance, our concepts of evidence are excluded, not because they are not concepts, or not science, but because they operate on a different level being about the evidence used to make substantive claims rather than the set of interlocking ideas that constitute the framework of understanding force, say. Other ideas are excluded simply because they are in everyday use. So we would not include time, or mass (in the sense of what it is rather than its physics definition) on the

grounds that asking someone to measure the time taken for something to happen is not likely to be problematic.

Let us take the second of our examples to illustrate what we mean when concepts *are* involved. We shall use one of the tasks that will be presented, in detail, later in which students were asked to find out how the slide of a shoe is affected by the surface on which the shoe was placed. How does this interaction between substantive and procedural understanding take place in this situation? It is worth noting here, and we shall return to this later, that this simple investigation in a context familiar to students from everyday experience, *could* be tackled with very little explicit use of substantive ideas; some students did not refer explicitly to the substantive ideas of force or friction and measured 'slide' by elevating a ramp until the shoe moved. However, the problem could have called on an interaction of both substantive and procedural ideas, which we illustrate here. Whatever, we argue here that students with an understanding of force will be advantaged regardless of the method they adopt.

1. To begin with, somebody has to 'notice' that the force is different between different surfaces. That, of itself, requires a piece of scientific observation related to friction which, in turn, requires that the event be viewed through a pair of conceptual/substantive spectacles. Thus, problem perception and generation call on substantive knowledge.
2. The next stage would be to decide how to measure the force. This is not straightforward. The problem may require reformulation. If the experimenter decides that the key variable is the starting friction, then a means of measuring the force required to get the shoe moving is needed. If, on the other hand, sliding friction is to be the focus, then the experimenter has to recognise that, provided the shoe is moving along the surface at constant speed, the force required is constant and equal to the frictional force. In this instance, substantive understanding is required before the dependent variable (a procedural idea) can be operationalised.
3. It could also be that note should be taken of the fact that the friction is only approximately unrelated to the speed.
4. But that it is dependent on the mass of, and in, the shoe. In procedural terms, mass will need to be identified and operationalised as a control variable
5. At this point, ideas about evidence already discussed come into play to plan and carry out the experiment, interpret and evaluate the data, using the ideas iteratively to find a solution; the validity and reliability of each of the inner three layers of the bullseye are in particular focus.
6. At the conclusion of the experiment, the substantive issues return as explanations of the behaviour of the shoe are formulated.

Thus, an investigation can be seen as a way of applying substantive ideas to solve unfamiliar problems, furthering understanding of the ideas. Substantive ideas can be the starting points for investigations as well as the end point.

Our first example is rather different, and again is one we use later in the research. It involves asking the question:

'Does the material of a 'helicopter' affect the time it takes to fall?'

Unlike our previous example, there are no hidden traps regarding the concept to be measured, although there may well be pitfalls ahead in terms of making the measurement. Letting the helicopter fall and measuring the time is not challenging of ideas about physics in anything like the same way as our first example. The difficulties here are all to do with the procedural elements of the task; how to measure the time effectively, how to cope with the large amount of variation between events.

Our third example is different again. It involves asking the question:

"How does the force experienced in a collision (of a car) depend on the design or size of a crumple zone?"

We might operationalise this question as meaning; what force is exerted when a weight (a car) with different length crumple zones (made of corrugated cardboard for instance) ‘crashes’. In which case, we need to do some sums:

$$Work\ done = Force \times Distance$$

and

$$Work\ done = -PE$$

Hence we can calculate the force exerted to stop the car from measurements of its initial potential energy (assuming the ‘car’ is dropped vertically), and the distance by which the crumple zone is distorted. Without such calculations it is not possible to make meaningful progress, so here substantive ideas are necessary, if not - we hypothesise - sufficient. We can summarise this by saying that in all cases we hypothesise that an (unknown) level of understanding of procedural ideas is a necessary condition for success in all of the tasks, but that a level of substantive understanding is useful in the second but necessary only in the third. This structure will guide our research into students’ ability to carry out investigations.

3.5.4 Working iteratively

So far we have a model for thinking about the ideas that sit behind investigative work (model 1), and a model for how those ideas interact with skills, facts and substantive understanding (model 2). We now need a model for how the investigation itself operates (model 3).

The APU (Gott and Murphy [1987] showed that an investigation is hardly ever a linear process. We attempt to summarise this in figure 3.4. Of course, there are occasions when an investigator plods through a linear pre-determined route, but they are a sub-set of investigations which, when new to the investigator, cannot be done successfully in such a routine fashion. Rather, we take the view that, at all stages of an investigation, the investigator must always be looking to what has been achieved so far, and amending and repeating to improve matters until such time as the solution is ‘good enough’ for the purpose in hand. Each iteration is made on the basis of understanding the relevant concepts of evidence; these are the ‘thinking behind the doing’.

The above is a brief summary of the position we have developed over a number of years and in a series of journal articles and books. We have brought the ideas together here, albeit briefly, to create a set of analysis tools for use later in this research. We do, however, need a final model to link the concepts of evidence and the investigation framework in models 1 to 3 with scientific literacy. This link will be made via the ideas of Toulmin and we shall review that in rather more details as it has not been used before in our work, although it has been put to use on other areas of educational research.

The purpose of this next section is centred on the making of a *claim*. All scientific experimentation results in such a claim, based we hope on the evidence from an experiment and on the previously obtained evidence of others working in the same area. Toulmin argues that the making of such a claim can be modelled as an argument, akin to that in a court of law, where the evidence for the claim is presented, dissected and its inadequacies pored over.

3.6 Toulmin and argumentation

The model developed here (model 4) relates to the process of moving from data to claim and vice versa; the model looks at how the claim can be constructed and deconstructed. To set the scene we shall review, briefly, a recent paper on this topic (Gott and Duggan [2007]) in which we used Toulmin’s (1958) ideas about argumentation. The purpose of this was to create a tentative structure within which it might be possible to see how these various ideas can interact and complement each other rather than falling into the typical education trap of assuming that one is to supplant the other, or exist in a state of conflict until there is a victor. Let us firstly consider the structure of an argument while conducting an investigation. The scientist collects evidence in a reliable and valid manner and makes claims as a consequence. Toulmin’s key terms which, he suggests, constitute the elements of rational argument are:

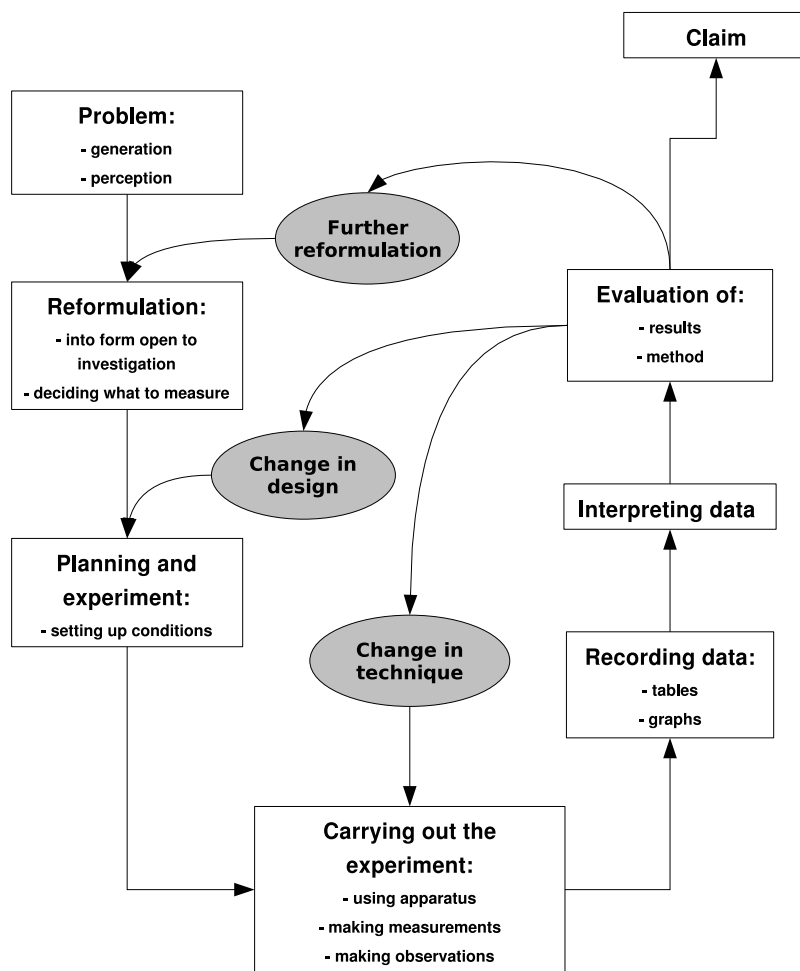


Figure 3.4: Model 3; An iterative approach

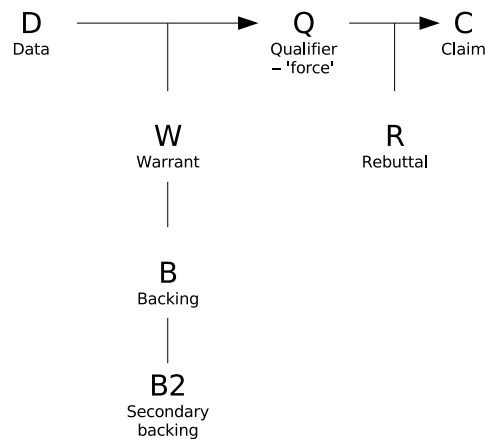


Figure 3.5: Model 4: Toulmin and argumentation

Data - the grounds or the facts which are the foundation for the ‘claim’. In the case of experimental work, and for our purpose here, we take this to mean empirical data in a ‘raw’, or summarised, state.

Claim - an assertion or conclusion from the data. This will be based on all the evidence (a term we expand on below); an analysis and interpretation of the raw data in a manner consistent with the warrants and backings and qualified appropriately. In an earlier article we distinguished between a private and a public claim, but in the interests of simplicity we shall not make that distinction here. The investigator will be working towards the production of a defensible claim. That claim may be made in response to a teacher or examination demand, or when it is defended in front of either teacher, or class or both. To the experienced scientist, the claim will be made in a conference presentation or journal article.

Warrant - the reason, rules or principle for justifying the connection between the data and the claim. In the context of school science, we are talking about the range of tried and tested methods of valid experimentation and subsequent data interpretation (the ‘case law’ of procedural understanding, underpinned by our concepts of evidence), as well as the substantive laws and principles, the complex network of interlocking theories that constitute the accepted body of scientific knowledge.

Backings - the detailed statements of fact upon which the warrant is based. In school investigations, it would be the detailed facts and ideas which underpin the data collection such as the number of readings taken, the method of averaging, the validity of the measurement itself, the interpretation of particular tables or graphs and so on. These are the ideas in the inner three layers of the bullseye, about the quality of the data itself and the design that has resulted in its collection.

Secondary backings - the questions a sceptic might ask from the outer layers of the bullseye such as: Is the experimenter competent? How do substantive ideas from others’ work affects the validity of the investigation? Was there bias in the design due to financial constraints or political or social influences?

Qualifier - a statement about the strength or ‘force’ of the claim. This could include, for instance, statistical reasoning such as ‘it is 95% likely that the claim follows from these data’.

Rebuttals - exceptions to the claim and/or conditions under which the claim will not be ‘true’. Here the claim is delimited to particular circumstances within which generalisable claims can be sustained.

We summarise this argument in figure 3.5.

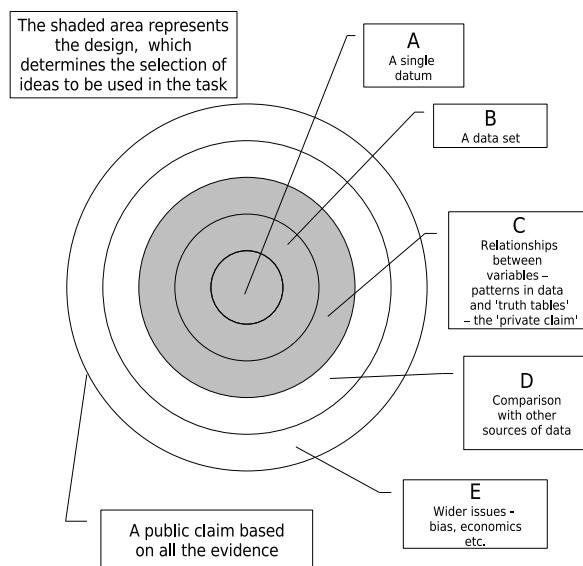


Figure 3.6: Model 1 expanded: concepts of evidence and claims

3.6.1 Evidence

A term not shown in figure 3.5 is 'evidence'. This is a somewhat broader and slipperier idea. We choose to use the term here to describe not only the data as defined above, but also the details of the research design, indeed everything that was seen by the investigator to be germane to the decisions taken in carrying out the task. So in figure 3.5, the evidence appears in data, backings and secondary backings, rebuttals and qualifier - all of those things which would allow us to form a judgement as to the quality and therefore utility of the claim.

We will now show how different concepts of evidence are used in the argument leading to, and defence of, different claims.

The innermost three layers (shaded) represent the ideas needed for a simple investigation typical of a school practical. An experimental design draws on whichever elements of A, B and C are deemed to be necessary to satisfactorily answer the question under investigation and it is predominantly the ideas within these three inner layers, the backings, that will constitute the argument resulting in the 'conclusion' or claim. Such claims are limited usually to the sort of context in which issues related to concepts of evidence in the outer layers are not always relevant. So we then skip these layers on the way to the claim in many simple school-based contexts. Scientists, of course, would similarly need to construct an argument to arrive at a claim drawing on A to C, but would also call on the concepts of evidence in the outer layers where appropriate and where the validity and reliability of the evidence may be affected by the broader context for the claim and its link to existing findings. The concepts of evidence in these outer layers act as secondary backings in the argument. Defence of the scientists' claim, using the Toulmin structure, would draw on all the elements in figure 3.5 including the secondary backings.

3.7 Looking forward and Looking back

So far we have considered how the concepts of evidence might be used as ideas used to make decisions in an investigation but science education must be about more than preparing future scientists. We also need to consider how these ideas might be useful to scientifically literate people. In contrast to an investigator constructing an argument that leads to a claim, someone looking at a scientists' claim needs to try to understand the argument that has led to it. In effect, they are looking back from the claim and of course, this is more complex, not least because the students have not designed the investigation(s) or collected the data themselves, figure 3.7. In judging the strength of the claim, students would need to 'look back' at the backings for the data; how the data were collected i.e. the experimental design, methods of measurement and empirical data collection.

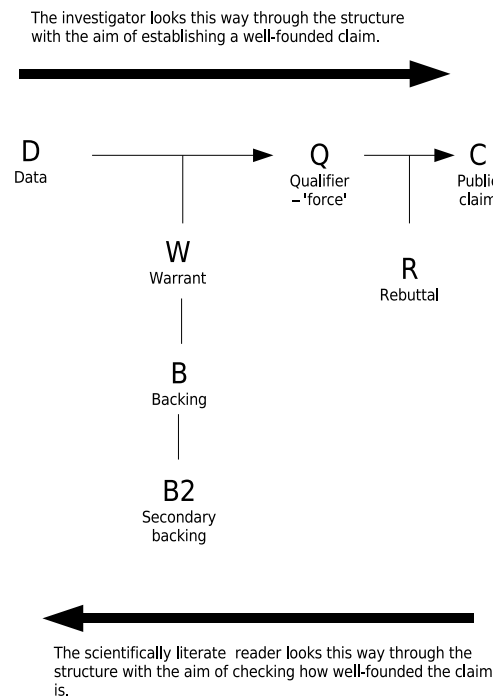


Figure 3.7: Model 4: Toulmin and ‘looking forward and back’

To do this, they have to understand that experimental design and empirical measurement are significant elements and need to be examined carefully, both for their inherent quality and for evidence of undisclosed bias of one sort or another. It is here that their own experience of collecting data comes into play. They now need to look back through the layers of the bullseye as best they can to attempt to disentangle what has been done and whether or not the claim can be defended therefrom, figure 3.8. The structure of the argument they use to create and defend their own claims becomes the structure they use to examine the issue. What, of course, is missing in their experience are the potential effects on the design from the outer layers of the bullseye.

In this research, we have found it useful to expand on the bullseye figure (figure 3.8), creating two bullseyes. The first (3.9) highlights the concepts of evidence of particular importance in much school practical work, ‘looking forward’ to be able to make a claim. These are the procedural ideas that contribute to conducting an investigation. This will culminate in a claim that the investigator feels is as good as can be managed with the resources to hand. The second (figure 3.10) illustrates how, when looking back on a claim in the public domain, the concepts of evidence required include all the important ideas about the experimental design and the data as well as ideas from the outer layers. The scientist must support and defend their claim to the public through the appropriate presentation and analysis of the data, also taking into account ideas in the outer layers. Someone looking back on this claim must consider how ideas from all layers of the bullseye have been used to construct the claim.

In figure 3.10 we have also subdivided the set of ‘wider issues’ (shown in figure 3.1) and have distinguished the concepts of evidence that enable us to consider how the validity and reliability might be affected by the investigators themselves from other economic and social considerations that might have influenced the evidence.

We can see from this tentative structure that ideas about evidence are common to both students conducting their own investigations and looking back behind others’ claims. We will discuss how this has informed the logical sequence of this research later. We have argued that in constructing an argument to make a claim from an investigation we work with the ideas from the inside of the bullseye outwards, ‘looking forward’ from the centre of the bullseye. [The overall ‘direction’ is outwards because of the need for data to argue from to make a claim.] From the perspective of a scientifically literate person where the aim is not necessarily being able to conduct investigations but to understand and evaluate the public

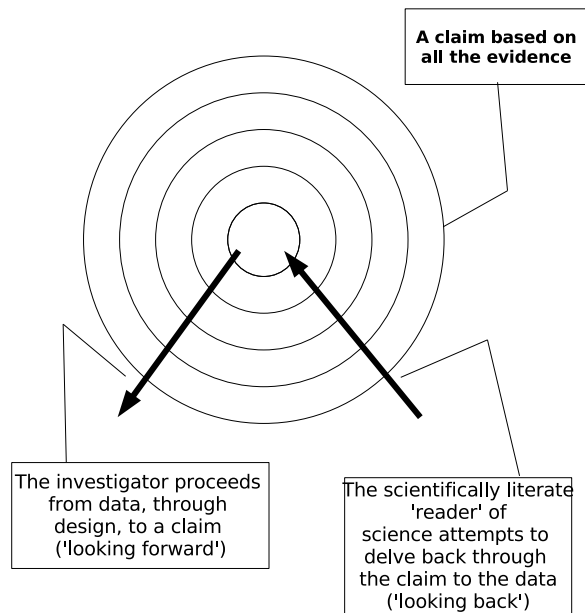


Figure 3.8: Model 1 and 'looking forward and looking back'

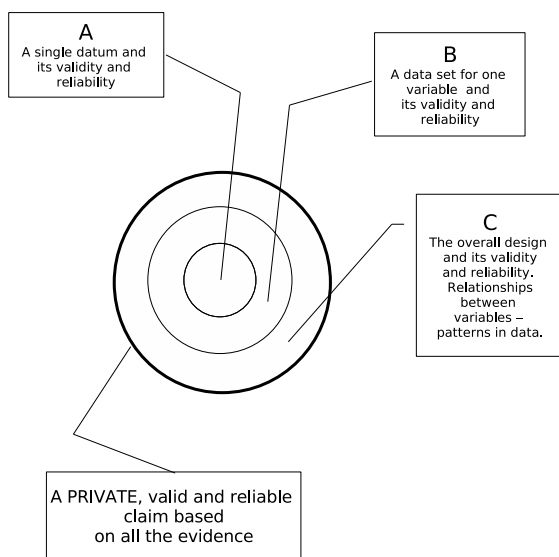


Figure 3.9: Model 1 and school practical work

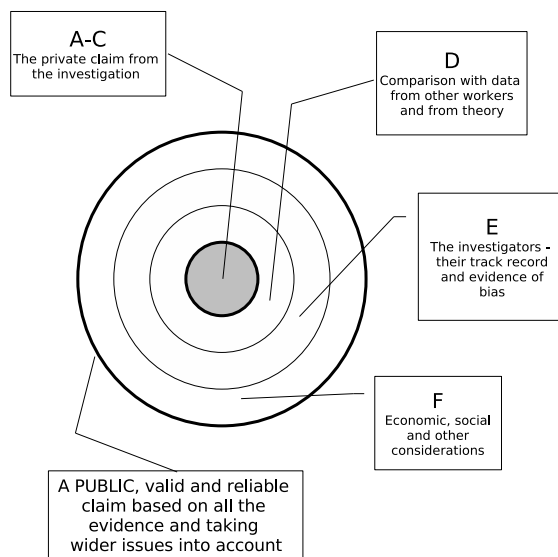


Figure 3.10: Model 1 and scientific literacy

claims made by scientists we have argued that in effect they are ‘looking back’ behind the claim resulting from someone else’s investigation (Gott and Duggan 2007) rather than just accepting the claim, figure 3.8. Such issues require us to work in the opposite direction; looking back from the claim to consider how the ideas in each layer can be used to determine the validity and reliability of the evidence that lead to the claim.

In many scientific issues in the public domain it is often that we have to infer the data and the design so it is often difficult to get back into the centre - although such ‘access issues’ need not prevent the scientifically literate reader from raising questions and being aware of the implications if answers are not available. But if data are available, and in a form which is not too difficult to come to terms with, then it should be possible to ‘look back’ from the claim into the investigation by considering the arguments which have created the claim and which constitute the forum for debate. To do that we need a structure which allows us to think about how claims are made and where data and design fit. It is here that the work of Toulmin in structuring the pattern or layout of rational argument is helpful.

We hypothesise that:

- * carrying out a good investigation will depend on an understanding of the concepts of evidence (as well as an understanding of the substantive ideas of science as necessary), organised through an understanding of the design of investigative tasks (looking forward)

- * and that being able to ask forensic questions about the evidence behind a public claim in a socio-scientific issue will depend on an understanding of the concepts of evidence and of design issues (looking back).

3.8 Concepts of evidence - investigations and claims

In the previous section we have shown that both ‘looking forward’ and ‘looking back’ draw on the concepts of evidence. Let us look in rather more detail now at the ideas in each layer. We shall break these ‘layers’ into two facets which are closely linked to the notion of ‘looking forward’, when conducting an investigation and ‘looking back’, when we consider others’ scientific claims. These are not hard and fast distinctions - they just point to different elements of the way in which ideas in the layers might be considered. In each layer, ideas in the forward looking facet ‘emphasising planning’ are more likely to be those first considered by an investigator when looking forward through the bullseye, to solve a problem and to make a claim. Ideas in the ‘emphasising analysis’ facet will also be used when looking forward to keep a constant eye on the quality of the evidence being produced, but probably following implementation of the ideas in the planning facet. This temporal sequence is, however, a simplification.

An investigator who is working iteratively will be constantly moving through the layers and the facets to evaluate the evidence as it is being collected.

A scientifically literate person who is trying to evaluate a public claim by looking back through the bullseye, may be more likely to think about the ‘emphasising analysis’ facet of any layer before trying to put themselves in the position of the investigator so that they can evaluate the ideas emphasising planning that must have been part of the original investigation.

The ‘layers’ and ‘facets’ are outlined below and should be seen as local iterations where a planned measurement, for instance, is then evaluated to see if it could and should be improved.

Layer A: A single datum In any empirical investigation something will need to be measured, quantitatively or qualitatively. That may be the area of a leaf, or the temperature of a solvent, or the speed of an object or a chemical reaction. This, innermost, layer is to do with the making of a single measurement of that variable.

A: Emphasising planning

In planning that measurement, thought must be given to the range and sensitivity of the instrument (the thermometer in the above example), and how the variable is derived from other measurements (distance and time in the case of velocity), how an accurate estimation can be made (the leaf area) or a colour change be determined (the chemical reaction). We must also be sure that the chosen instrument and method do, in fact, measure what we want in a valid way; the maximum length of a leaf may well be a reliable measurement but is not a valid one (unless we know that area and length are directly proportional); the colour change associated with identifying a substance will not be valid if the test is not specific to only that substance.

A: Emphasising analysis

When the measurement has been made, we shall wish to have some idea of its accuracy. This can be estimated by careful consideration of the instrumental accuracies. In some instances, where only one measurement is possible, this is the only way to determine its accuracy. But, of course, one measurement is not always enough. It might be in the cases quoted above, but a measurement of blood pressure, no matter how carefully done, is not reliable and repeats will be necessary.

Layer B: A data set We are led, then, into the next layer which is to do with repeated measurements of the same variable under the same conditions.

B: Emphasising planning

In the planning stage, that requires trial and error repeats to establish the extent of the variation between successive readings. Which, in turn, allows for a judgement to be made as to the number of repeats necessary to establish a certain confidence level (via the standard deviation of the mean - ‘the standard error’- of the measurements if there are sufficient to carry out the statistics, or the application of common sense if not).

B: Emphasising analysis

The analysis of the resulting data set will confirm the likely error, due to random variation on the mean (by calculation of the mean and standard deviation of that mean), and then possibly amending the decision on the number of readings to reduce the standard deviation of the mean.

Layer C: Design and Relationships As we move out, we arrive at the crux of the investigation; the establishing, or otherwise, of a relationship between one or more of the variables. This layer really permeates the whole task, but it is hard to show that in a simple diagram. At the planning stage, the design may be somewhat vaguely expressed. That vagueness is then removed as details of the inner layers are sorted out in conjunction with decisions about factors that might affect the relationship, the control variables - the design becomes operationalised. So this layer should be seen as one that is continually revisited as the design is sharpened up.

C: Emphasising planning

At the planning stage, we need to think about the identification and value of potential control variables and the range of the independent variable which will be needed to establish any potential relationship. We also need to worry about the interval between such readings if we are to catch any oddities - maxima or minima perhaps - in the behaviour.

C: Emphasising analysis

When those data are analysed, we will need to have some sense of the range of errors and variation in all of the readings to feel sure that any line, or curve, is really there. At the analysis stage the concern is mainly focused on the reliability and validity of the complete task. Does the design allow for the question to be answered? And were the data and resulting patterns also valid and reliable.

Layer D: Comparison with other data and theory Questions which will have been to the forefront of the investigation, particularly in its planning stages, relate to the existence, or not, of other similar research and data and how this may have informed the investigation.

D: Emphasising planning

At the planning stage this will involve a check on the existence of other such research and relevant theoretical structures.

D: Emphasising analysis

During the analysis, the scientist would be anxious to ensure that the data was not at odds with that from other research, nor at variance with the expectations of the substantive theory. (We are, of course, assuming here that we are not at the cutting edge of knowledge).

Layer E: The investigators Increasingly, the investigators themselves are placed under the microscope. Typically they will need to defend their track record and divulge their sources of funding so that the ‘public’ can judge their qualifications and the extent of any vested interest. This is particularly true of medical research with its history of big business involvement, from the infamous tobacco scandals to insidious manipulation by (some) drug companies.

E: Emphasising planning

At the planning stage, the scientists should be aware of possible bias in the design of their investigation through, for instance and in the case we will be considering in our research, the choice of places in which to sample.

E: Emphasising analysis

At the analysis stage, the concern will be to establish their independence.

Layer F: Economic and social pressures And finally, there is the need to set everything in a social, political and economic context; factors that pervade and might influence all the stages leading to the claim.

F: Emphasising planning

The form of any investigation will be shaped by the available funding, inevitably, but also by social, environmental and business pressures. That shaping should be acknowledged in the design.

F: Emphasising analysis

The analysis of the data and its presentation should also point to those pressures and moderate the public claim accordingly.

We do not wish to imply here that every practical science task must cover all these layers, nor that these layers cover all eventualities. Sometimes a single, accurate and defensible, measurement is all that is needed; a level of alcohol in the blood for instance. At other times the measurement is no more than counting - the number of daisy plants in a quadrat. And at others the measurement is qualitative; the presence or otherwise of carbon dioxide from a reaction. In these cases, only some layers are called into action, or the emphasis lies predominantly within just one of them. In this research, when we look at students’ investigations we shall use the structure (A to C) of the first of these diagrams figure 3.9. When we turn to the issue task, we shall need to make use of A to F shown in the second diagram (figure 3.10).

3.9 Putting the models together

So far we have described the concepts of evidence, models for investigative work and argumentation and how they can be used:

- ‘looking forward’, to structure an investigation and ultimately to defend a claim and
- ‘looking back’ to question the evidence-base for a claim.

In this section we will consider these ideas in the context of socio-scientific issues, where a scientific claim and the argument based on the concepts of evidence, are just *one* of the factors that are taken into consideration when debating the issue and, possibly, reaching an agreed position. We have represented this as the right hand box on the following figure 3.11.

We will now locate many of the elements from the arguments above on the diagram (model 5) and explain them. Let us look at each of the boxes on the diagram in turn. We have already described the structure of the left and middle boxes; we bring the points together here for clarity.

The left hand box is the 'doing' bit of an investigation - the planning, iterations and 'thinking behind the doing' that characterises actually carrying out an investigation, always, of course, with a view to making a defensible claim. The concepts of evidence are of central importance at this stage, together with the way of working (iteratively or not). Models 1, 2 and 3 are involved here.

The middle box is where the argument is put together (model 4). The deliberations in the left hand box now become fodder for the argumentation model. The raw data of course, but also the design issues are now the raw material used in constructing the claim.

The right hand box is where the claims from this and other experiments and 'non-science' sources of knowledge are put together into a forum where yet another argument or debate is to be had. The grounds for debate include not only the claim from the investigation but also other relevant factors. Many scientific claims are related to issues in the public domain. The 'public debate', whether in the context of a planning enquiry, a personal decision or debate in the media, must take account not only of the scientific claims from the investigation - the validity and reliability of which must be established (using ideas depicted in the left and middle boxes on the diagram) - but also other issues that must be taken into account before an agreed position is reached. Thus evaluating the claim may go on as part of the public debate (shown in the right hand box) but draws on ideas in the left and middle boxes. The rest of the public debate weighs up the status of the claim with other factors to reach an 'agreed position' on the issue. Of course, a 'controversial issue' is one whereby an agreed position has not yet been reached. In the case of global warming, for instance, the status of the claims from investigations are hard to evaluate so the public debate (in the right hand box) is about the degree of confidence in the claims and this is then weighed up with, for example, economic and political factors. So an 'agreed position' on global warming is hard to come to and the public debate continues. We note at this point, and expand on this later, that different conceptions of 'scientific literacy' encompass varying elements of this model. Our view of scientific literacy, whereby an individual is able to question the scientific basis for evidence-based claims in the public domain, encompasses the ideas in the left hand and middle boxes and their contribution to the public debate. It excludes the non-scientific and un-scientific claims, other than to note that they also contribute to any debate to reach an agreed position. While the ideas represented in the right hand box are largely not 'science ideas' they draw attention to factors, each with its own knowledge base, that contribute to the public debate about scientific issues.

3.10 An example

To simplify matters, let us consider the straightforward task of measuring a particular form of pollution in a river in response, say, to local concerns about the effect on fish stocks. Decisions would need to be taken about the sensitivity and specificity of the instruments to be used to measure the pollution, the place and time of any sampling, and the number of such samples and how the resulting data is to be analysed and interpreted. So far we are using ideas in layers A to C of the bullseye (the primary backings) - familiar ground. But the outer layers (the secondary backings) may also affect the design. For instance, the most appropriate measuring instrument may simply not be available for reasons of cost. Or an inadequate theoretical basis may result in sampling at inappropriate places or times. It could even be the case that the funders of the research have influenced the design, explicitly or implicitly. What we are arguing here is that 'best practice' science would deal with all these issues in making its claim. Such a claim would, therefore, give full details of such matters allowing the reader to weigh the claim as evidence in the decision-making process.

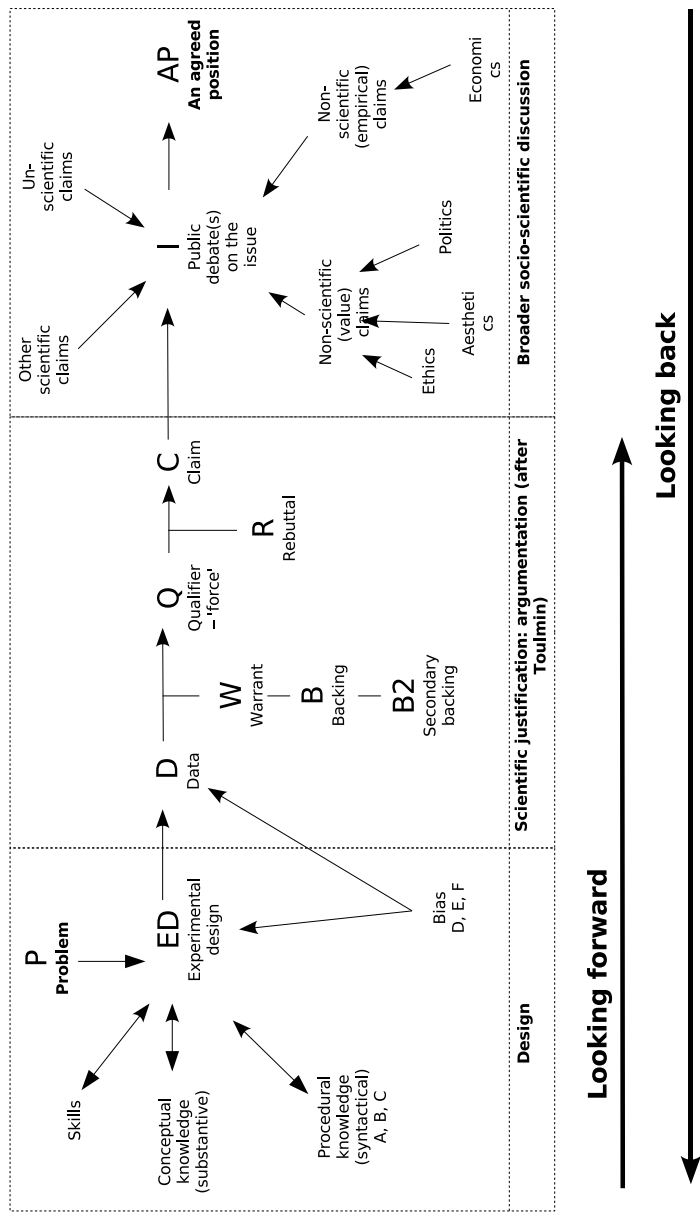


Figure 3.11: Model 5: Empirical work, argumentation and scientific literacy

Thus far we are in the left hand and middle boxes of figure 3.11. Let us now turn our attention to the forum within which ‘real’ decisions are taken to arrive at an agreed position. And, again by way of example, let us consider a planning enquiry as constituting such a forum. The purpose of such an event can be seen as:

- examining the scientific evidence, as we have already described, and coming to a judgement as to how much it can be trusted.
- placing that evidence into the arena of the political, social and economic decision-making where many other factors are brought to bear.

This second stage is represented in the right hand box of figure 3.11. Here the data from other walks of life enter the field. This may include such data as the cost in lost (or gained) jobs, the cost of renewed transport infrastructure and so on. We can envisage now a re-run of the central Toulmin box where the data are now expanded in this way resulting in another evidence-based claim.

We need also to add to this other non-evidential matters. The aesthetics may matter to many people - a dead river, in our example, is more than simply a lost economic opportunity. Some groups may have strongly held views which fly in the face of the evidence, but this, as we know to our cost, does not mean that such views will not prevail. All the evidence is that ‘gut feelings’ are usually the way decisions are taken. And it is here that the rational basis for decision-making may face defeat. What we are advocating here is not that decision-making should be reduced to a dry and politically sterile argument over the data, but that the interested participants will be able to evaluate the evidence and decide to ignore it, if they so wish.

3.11 The theoretical structure as applied to the research

The research described in this report can be located within the theoretical framework we have established. We have been interested to explore:

1. Students’ understanding of the concepts of evidence; the ideas located in the left hand box.
2. Students’ ability to investigate; looking forward through the concepts of evidence and working practically to make a claim, as shown in the left hand and middle boxes.
3. And students’ ability to look back on a claim made in a socio-scientific issue: a critical form of scientific literacy shown from the right hand box back to the left.

3.12 Some additional comments

3.12.1 Under-specification

Finally, we need to introduce a concept that runs through our thinking and our analysis of curricula; the idea of an under-specified curriculum which derives, we suggest, from the lack of clarity in the literature we noted at the beginning of this chapter. The concepts of evidence act as a detailed domain specification - the basis for selecting the ideas that could be taught and assessed. By attempting to articulate the ideas that are required for a procedural understanding we aim to avoid the potential for ambiguity inherent in more general terms.

We have found, from experience, that a clear specification of the ideas required for a procedural understanding acts as an ‘aide memoire’ when teaching and assessing; it describes clearly what is to be taught and assessed (although obviously does not imply *how* ideas might be taught or assessed). A clear specification of the ideas keeps the underpinning knowledge base of procedural understanding at the forefront of people’s thinking. For instance, many of the detailed concepts of evidence in layer B (above) are the underpinning ideas required to understand ‘variation in repeated readings’. We have found, through personal experience, that in an under-specified curriculum - one without such detail - teachers and examiners, who may be less familiar with articulating the underpinning procedural ideas than they would be for an equivalent substantive understanding, tend to focus on only a few aspects

of repeated readings that then may become ritualised, such as anomalous readings and the number of repeats.

In previous publications, we have described curricula that do not clearly articulate the ideas that constitute the understanding they aim to develop, as being ‘under-specified’ (Gott and Johnson [1999]), where we point to the dangers of under-specification of both substantive and procedural ideas). Under-specification can lead to differences in meaning - between researchers and when curricula are enacted. Thus, the ‘skills approach’, (typified earlier) may or may not have been the intended outcome of the curriculum developers. Under-specification in the curriculum may have distorted their intentions. In summary, the consequences of an under-specified curriculum, with respect to procedural understanding, may account for some of the positions that we identify in the literature and in this research.

3.12.2 What we do not include in our structure

Much of the debate in this area of science education revolves around terms such as ‘process’, ‘enquiry’, the nature of science and so on. The first two are included in the above structure, but the last is not and this omission is deserving of comment. The nature of science has come to be associated, in science education at least, with such things as the understanding of the difference between laws and theories, or the uncertain nature of scientific knowledge. We take the view here that these ideas, whilst inherently interesting, should not find a place in science education in school years, except peripherally.

Why do we take that view? The use of terms such as laws and theories is commonplace in science, but the philosophical meaning of such terms is not. For us, it is the understanding of the laws and theories which matters, not understanding the labels that are attached to them. As so often happens in education research, the labels take on an importance far greater than they warrant and, in the extreme, *become* that which is to be understood. Turning to the uncertain nature of science, we come across some major issues tied loosely to post-modernism which we have no time, or indeed energy, to delve into here. Of course scientific ideas change and adapt. But the notion peddled in some post-modern writing that the fact that Newton’s laws fall apart at the atomic level renders them uncertain when describing the motion of a planet is plainly nonsense. Some go even further and argue that such knowledge should be treated as relative and that it is possible to hold different views. The fact that any piece of science has to be tested against the behaviour of the real world and that if it doesn’t explain that behaviour in a logical fashion then it is not useful cannot, however, be avoided. In our view, the science needed for schools is well-known and not subject to this sort of uncertainty.

But there is an uncertainty which is critical. Let us take global warming as an example. Predictions for global warming are very difficult to make. We are predicting a long way ahead from a relatively limited time span of existing data - at least of the accuracy required for such predictions. Extrapolation is always a dangerous pursuit, amplifying as it does any uncertainty in the slope of the graph. But there is more. The model used to make the prediction is hugely complex and we make no pretence here of understanding it. However, sea temperature seems to be a significant feature in the model, and changes in that temperature of fractions of a degree are critical to the direction the extrapolation goes. Imagine the task - measuring sea temperature to, say, 0.1 of a degree. Which sea? what depth?, what time of day? what time of year? how to average across all seas? This is a seriously difficult job. Here is where we run into the issue of uncertainty. But the uncertainty is not of the theory, but of the data used in the model. And this second view of the nature of science is the one we wish to address within the framework described here. It seems to us to be of critical importance that the public comes to understand that such uncertainties are inherent in the measurement process that lies at the heart of science, not that they represent a failure of science, or of scientists.

Chapter 4

Methodology

4.1 The sample

Data were obtained via a number of research instruments, detailed in a subsequent section, from students in Years 1 and 2 of the course over two academic years. The 2004 cohort were in their 2nd year when data were collected. The 2005 cohort were involved at the end of their first year. Further data were collected from them in their second year when we repeated the sequence of data collection. As we noted in an earlier section, students in year 1 undertook a compulsory science module which deals with substantive ideas in physics and chemistry. Year 2 students took the module of which the ‘evidence’ section comprised some 9 weeks (of 2.5 hour plus ‘homework’ per week). We were unable, for ethical and logistic reasons, to split this group to create a control group or to use any instruments prior to their beginning the module. We used Year 1 (2005 cohort) students as a control group for one of the instruments (the issue task) and this group then went on to complete the issue task after their teaching programme in exactly the same way as cohort 04. A further cohort (2006) were the subject of an intervention aimed at teaching argumentation. Their module structure was the same as earlier cohorts except for the addition of directed teaching on argumentation based on the Toulmin structure outlined in the theory chapter. They are not shown on the diagram to avoid clutter, but followed the same pattern as the 2005 cohort.

4.2 The instruments

The various instruments and the point in the course at which they were used is summarised in figure 4.1.

A note about the research instruments Let us assume, for the time being, that the teaching programme is very effective at developing an understanding of the ideas that we have targeted, although of course some will be better at them than others. We then ask them to complete the evidence test, carry out an investigation and raise questions in response to the local issue. Let us further suppose that some students do not do the evidence test very well. Can we therefore assume that they are students who have not understood the ideas as well as others? The answer of course is no. Whilst the evidence test, as with all the instruments, relies (we argue) on an understanding of the ideas, such an understanding is likely to be a necessary rather than a sufficient condition for doing well, and other factors must also be considered. The evidence test is traditional in the sense that the students, particularly younger ones, will be familiar with tests of this kind which they have been subjected to ad nauseam in the UK education system. They will be familiar with exam atmospheres. But they might freeze when faced with an open ended task which their gradgrind schooling has ill-fitted them for. Older students may be less worried by this, but also less at home with the traditional exam. So other factors influence performance on the tasks, and those influences will be different on different task types (described in more detail later).

4.2.1 Biographical information

The questionnaire deals with basic biographical information used, in part, to check the University records. As well as collecting data on students’ age and gender, the questionnaire asked students to report

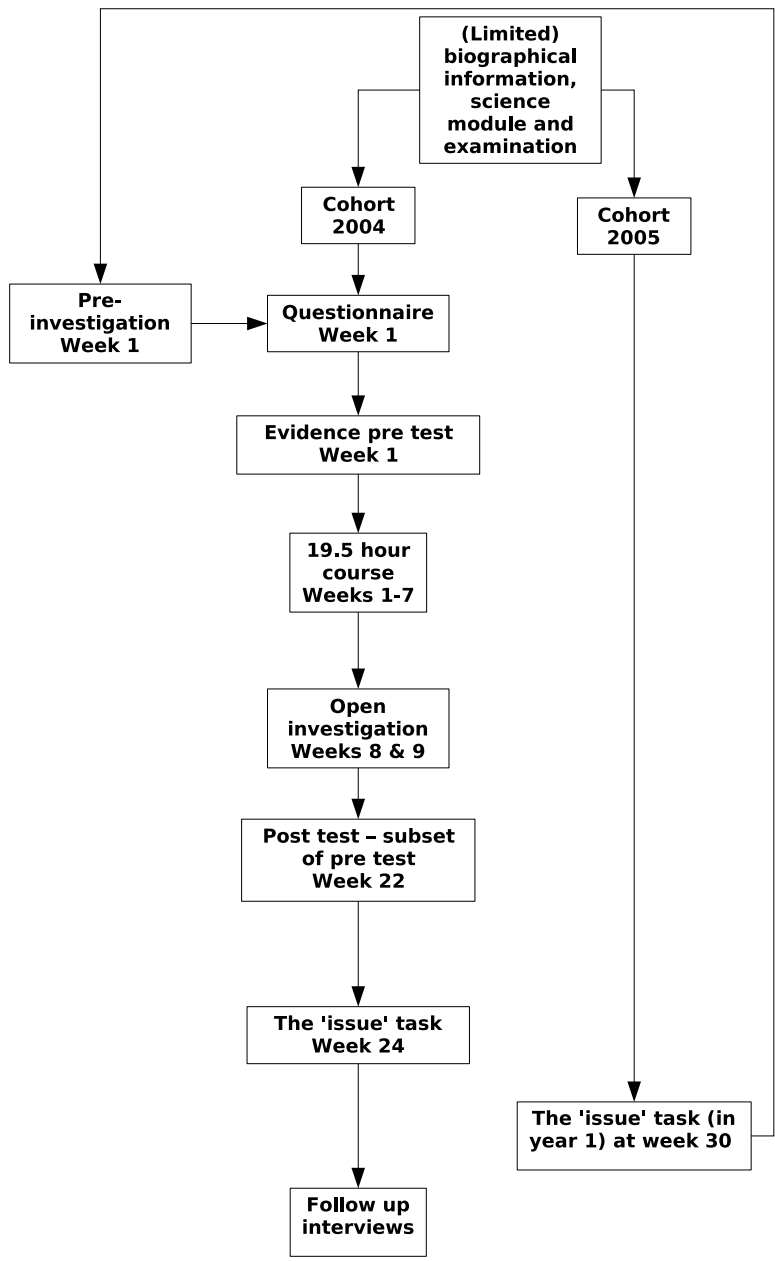


Figure 4.1: An overview of the research

their previous qualifications in science and maths at GCSE level (or equivalent) and their subsequent qualifications prior to starting the BA (A level, BTEC, etc.). In addition we asked them to indicate whether they were in the 1st generation of their family to attend university and, to be able to determine their family's socio-economic status (SES) using the ISCO88 classification scheme (International Labour Organisation (ILO) 1990: ISCO-88: International standard classification of occupations. Geneva: ILO), their mother and father's (or adult carer's) current or previous occupation.

4.2.2 The evidence pre-test and post-test

At the same time as the questionnaire, students were asked to complete a 'test' of their ideas about evidence prior to the course. The test targets such things as measurement, experimental design and data analysis and is reproduced among the appendices. The pre-test, which took up to an hour for some students to complete, comprised some 17 items spanning the concepts of evidence. Although set within biology, chemistry or physics contexts, the questions required minimal understanding of the substantive ideas. We have used this test elsewhere (Roberts and Gott [2003, 2006b]) and it has been refined as a consequence. A handful of the sub-items were dropped from subsequent analysis as a result of the item analysis.

At the end of the course (week 9) students were asked to complete a post-test. This was a subset of the pre-test (time did not allow for a full repeat of the pre-test) chosen on the basis of a combination of factors.

- Facility and discrimination: avoiding terms with high facilities which would be likely to run out of headroom
- Richness: items which, on the pre-test, gave interesting responses
- Spread: across the various concepts of evidence

The common tasks were 4, 5, 8, 9, 11, 16 & 17 from the evidence pre-test.

Pre- and post-test were scored using the same coding system. A sample of 10 tests was marked independently by the authors and inter-marker checks showed a consistent application of the mark scheme.

4.2.3 The open-ended investigation - cohort 2004

We shall make use here of the students' written accounts as a means of assessing their ability to carry out an investigation. Our experience is that this written material is a reasonably good representation of what they did. There is research evidence to back up this point (Baxter et al. [1992], Gott and Murphy [1987], Welford et al. [1985], Haigh [1999]) but it is not entirely conclusive and we must, therefore, issue a minor health warning at this point. The account of the investigation was part of the summative assessment for the module.

Students were given the following guidance in which they were asked, in their report:

- Demonstrate a critical understanding of the factors that determine the quality of scientific evidence and in particular how they impinge on its reliability and validity.
- State what can be concluded from your data.
- Produce a coherent word processed report which sets out the main issues clearly and includes all data collected, including trial runs.

Students were encouraged to write an account that reflected the 'thinking behind the doing', making explicit the decisions they made throughout the investigation in the form of a dairy of events rather than as a formalised account. In weeks 8 & 9 students were required, as part of the formal assessment process for the module, to carry out an open-ended task. The task (for the 2004 cohort) was:

'Does the material of a 'helicopter' affect the time it takes to fall?'

This was set in the context of ‘spinners’, familiar to all the students from their primary school days and much used in current primary science programmes. No further help was given except on matters of fact not related to the students’ own data or experimental design. A wide range of apparatus was made available: wide enough to allow for a variety of approaches to the task.

The students submitted a written account at week 14 (after the course ended at week 10). The investigation was marked in two ways, conditioned in part by the requirements of the University module assessment. The first ‘score’, labeled ‘TotalPC’, is an aggregate mark. The basis for the marking system is tied to the logical structure of the task itself. So, marks are awarded for design, measurement, data presentation and analysis and summed into the final mark. This mark was the ‘formal’ assessment mark and is also the one used in the regression analyses which follow. After the initial marking of the reports, one of us (RG) second marked a sample to check on the reliability of both coding and marking. This proved to be satisfactory. The second ‘mark’ was of a different nature.

Following repeated observation of students at work on such investigations over the past few years, it gradually became apparent that they could be crudely grouped into 3. The three approaches are characterised below:

Divergent: In this approach we see students collecting data industriously. They make many measurements of many variables but it is clear that they are doing it in hope rather than expectation. It is impossible to discern any plan to this scatter-gun approach and, indeed, their subsequent accounts illustrate all too clearly the lack of such a plan. They seem to believe that, if they collect enough data, ‘something will turn up’.

Linear: Other students make a plan and are quite determined to see it through to the bitter end. Some make lots of measurements again, but this time there is a clear plan, often seen as unalterable in the face of the accumulating evidence. The problem is that, when the plan goes wrong, they carry on anyway. As they say of war, the battle plan is the first casualty. They are unable to take stock if the position they find themselves in and, when the data proves inadequate, their surprise is unalloyed.

Iterative: This group, the ‘best’ group, is characterised by a clear attempt (not always successful of course) to modify their battle plan in the light of tactical defects. They will try to repeat measurements only until they are satisfied that they have enough. They will do trial runs, keep ongoing graphs of their data, and be prepared to change course and even abandon their plans altogether in the face of problems or data which isn’t ‘working’. They assume nothing, everything being determined empirically.

This subjective allocation was coded separately. Two further codes were found to be necessary. ‘R’ was used to indicate that, although the approach was iterative, it seemed to be a rather routine ‘teacher-knows-best’ form of iteration. However, where the data required it, this category was collapsed into the Iterative group. Similarly a group labeled ‘T’ showed clear evidence of iterative behaviour in their trials runs, but did not then follow that through very well. Again, and where necessary, this was collapsed into the Iterative category.

Approximately 60% of the marks were allocated to the students’ explicit understanding and application of the concepts of evidence in the account and 15% for the coherence of the whole account. In addition, 25% of the marks were awarded for their way of working (divergent, linear or iterative). The total score (TotalPC) we judged to be the best reflection of the overall quality of the investigation. The coding into iterative etc. will be used in some analyses whilst part scores will also be used which reflected the extent to which they showed understanding of *individual* concepts of evidence.

4.2.4 The open-ended investigation - cohort 2005

For logistical reasons linked to the University assessment arrangements, it was not possible to re-run the ‘helicopter’ investigation with the following year’s cohort (2005). Instead a well-tryed substitute was used. It is detailed below. This was used as both the pre and post-investigation.

Instruction sheet given to students:

Science for primary teaching 2

The investigation. ‘Sliding shoes’

How much do different surfaces affect how easy it is for a shoe to slide?

You are required to carry out and write about the investigation, using equipment available in the labs.

Please attend the lab for as long as you need to carry out the investigation.

NB. CARE must be taken if using heavy masses to ensure that safe practice is adopted.

You need to work individually when collecting data, using equipment available to you, and must submit an individual report on the investigation. Your account must be word processed and should include tables and graphs, preferably processed in a package such as Excel, but graphs drawn by hand are acceptable. This work is to be submitted to your tutor in the seminar sessions. Please put your name on it. We will provide you with feedback on it.

In the investigation you should:

- Demonstrate a critical understanding of the factors that determine the quality of scientific evidence and in particular how they impinge on its reliability and validity.
- State what can be concluded from your data.
- Produce a coherent word processed report which sets out the main issues clearly, makes explicit the ‘thinking behind the doing’ and includes all data collected, including trial runs.

To be able to do this you have to demonstrate your understanding in your written account.

Your account must therefore make explicit the ‘thinking behind the doing’. It should not only clearly and unambiguously indicate what you did and what you found out, but should also reflect your way of working, the decisions you made and the ideas that you considered which lead to those decisions. The best way for you to write explicitly about the things you are thinking about when you are doing your investigation is to report as if you are ‘thinking out loud’. Please report exactly what you do in your investigation and tell us the reasoning behind it. This helps us to better understand your ideas and to diagnose issues. Your account should not look like a traditional account of a science experiment, with just ‘apparatus, method, result, conclusion’, but should reflect your use of the ideas which enabled you to do what you did.

You are advised to note down these thoughts and decisions as you work in the lab.

4.2.5 The transfer task

The purpose of this task was somewhat different. We have argued elsewhere (Gott and Duggan [2007]) that an understanding of concepts of evidence, and practice at their deployment in practical tasks, should enable students to examine an issue of general importance through a new lens. We hypothesised that they would ask qualitatively different questions relating to the data, and the design that delivered that data, as a result of their newly acquired understanding.

The students were given an introduction, using a PowerPoint presentation, which provided some background to the real-life context. They were shown a photo of the cement works used in Tytler et al.’s work (Tytler et al. [2001a], Tytler et al. [2001b]). They were provided with the following information which summarised the issue:

- The cement works is sited in an area of ‘outstanding natural beauty’
- The cement works changed the fuel burnt in the chimney
- Residents were concerned that burning the new recycled liquid fuel might damage their health and the environment

Table 4.1: The sample

Cohort	Male	Female	Total
05	15	82	96
04	8	78	87
Total	23	160	183

Table 4.2: The age structure by year

Age	Cohort 05	Cohort 04
<20	42	25
20-25	29	32
25-30	9	5
30-35	8	10
35-40	6	8
40-45	3	6
Mean	23.8	25.6

- The company agreed to call in experts to evaluate the potential pollution
- They reported that the chimney was safe.

They were then asked to fill in the first of the question sheets:

The experts reported that the chimney was safe. Imagine you are a local resident, discussing this. What would your first thoughts be about this?

This first sheet was to capture their initial thoughts without any specific prompting. They were then given a separate sheet, the first one having been collected:

The residents wanted to ask the experts lots of questions. If you were to write a letter on their behalf, what specific questions would you want answered?

The second sheet gave greater guidance. Students were told they could repeat any points they might have already made on the first sheet.

The answer sheets were coded by a process of scanning each response and adding any significant factor to a growing list which were then coded as present/absent. No attempt was made to discover the extent to which the students' questions represented an *understanding* of the idea, merely that the question was asked in a context which was, on the face of it sensible.

4.3 The sample described

The data below were used to generate variables for regression and QCA analysis. However, few of them were significant in either type of analysis. The data are included here, therefore, to give as broad a picture of the sample as possible, allowing the reader to decide to what extent the results are generalisable, and in what circumstances.

4.3.1 Cohorts 04 and 05

We have limited biographical data for both cohorts. Given that we intend to make comparisons between the groups, we must make what comparisons we can between cohort 05 and cohort 04 so that the reader can make his or her own judgment as to their comparability or otherwise. We will indicate in any data whether the comparisons are within or between cohorts.

The following tables give details of gender and age, tables 4.1 and 4.2.

The age variable was dichotomised at age 23 (students' age at the start of the evidence module). Students of this age and above would all have had at least 3 years' break from the routinised assessment-driven school or college-based education, either immediately prior to starting the course or between

Table 4.3: School types attended

Type of school	Cohort 04	Cohort 05	Both cohorts
School	10 (11.6%) ¹	18 (18.6%)	28 (15.3%)
Sixth form college	32 (37.2%)	43 (44.3%)	75 (41.0%)
FE college	30 (34.9%)	22 (22.7%)	52 (28/4%)
Not known	14 (16.3%)	14 (14.4%)	28 (15.3%)

Table 4.4: Number of A levels

Number of A-levels	Cohort 04	Cohort 05	Both cohorts
None	29 (33.7%) ²	22 (22.7%)	51 (27.9%)
1 or 2	14 (16.3%)	5 (5.2%)	19 (10.4%)
3 or more	32 (37.2%)	59 (60.8%)	91 (49.7%)
Not known	11 (12.8%)	11 (11.3%)	22 (12.0%)

leaving school and gaining an A-level equivalent qualification. They will have experience of life outside a school system and thus may be more inclined to question ‘accepted wisdom’, recognise complexity and uncertainty and have had the opportunity to participate in problem solving activities and engage in the democratic process. Students above the age of 23 will have all made the positive choice to enter higher education, (rather than going there by default as an extension of schooling). Thus they may place a greater value on learning and may be prepared to put more effort into their work. Accordingly, those below that age, labeled ‘young’ (N=118) can be seen as traditional students coming to University straight from school. Those aged 23 or over (N=65) are labeled ‘old’ and have at least some experience outside the education system.

Table 4.3 gives details of the type of school where the students obtained their education after the age of 16.

We note that the number of students without any A-levels or with fewer than 3 is not insignificant (see table 4.4). Comparing A level results between cohort 04 and cohort 05, it is clear that, for whatever reason, the cohort 05 students are better qualified.

For both cohorts of students we have additional information from the biographical questionnaire. In table 4.5 we see that older students are far more likely to come to the course through access routes which is hardly surprising. By contrast, older students are no more likely to be the first ones in their families to go to university than younger students (table 4.6).

4.3.2 Social class and socio-economic status

All students were asked about their parents’ occupation. The resulting data was examined closely and any ambiguities were dealt with by interviewing the students concerned. We then allocated both parent

Table 4.5: Age and access route

Cohort 04		
Access		
Age	No	Yes
<23	35 (77.8%) ³	10 (22.2%)
>=23	11 (36.7%)	19 (63.3%)
	46 (61.3%)	29 (38.7%)
Cohort 05		
Access		
Age	No	Yes
<23	57 (90.5%)	6 (9.5%)
>=23	7 (30.4%)	16 (69.6%)
	64 (74.4%)	22 (25.6%)

Table 4.6: Age and first generation university

Cohort 04		
First generation		
Age	No	Yes
<23	19 (43.2%) ⁴	25 (56.8%)
>=23	10 (31.3%)	22 (68.8%)
	29 (38.2%)	47 (61.8%)
Cohort 05		
First generation		
Age	No	Yes
<23	17 (26.6%)	47 (73.4%)
>=23	9 (33.3%)	18 (66.7%)
	26 (28.6%)	65 (71.4%)

Table 4.7: The ISCO classification of occupations

ISCO CODE	Description - major categories	N (both cohorts)
1000	LEGISLATORS, SENIOR OFFICIALS & MANAGERS	41
2000	PROFESSIONALS	50
3000	TECHNICIANS AND ASSOCIATE PROFESSIONALS	29
4000	OFFICE CLERKS	15
5000	SERVICE WORKERS & SHOP & MARKET SALES WORKERS	19
6000	SKILLED AGRICULTURAL & FISHERY WORKERS	1
7000	CRAFT ETC TRADES WORKERS	8
8000	PLANT & MACHINE OPERATORS & ASSEMBLERS	4
9000	SALES & SERVICES ELEMENTARY OCCUPATIONS	3
	NA	13
Total		183

to an ISCO classification which has 9 major categories and many sub-categories located in a 4 digit code. The numbers in each of the major categories are shown in table 4.7.

For the purposes of other analyses we then transformed the ISCO coding, using a well-trying algorithm (Ganzeboom and Treiman [2005]) into two further variables. The first is the International Socio-economic Index which is designed as an interval measure. The distribution of that variable is shown in figure 4.2.

The second transformation was from the ISCO code to the Goldthorpe classification. The frequencies are shown in table 4.8 (categories 4 to 7 are absent as they require information pertaining to the number of employees that is not available from the ISCO codings alone).

A cross tabulation of the classification against the ISCO categories is shown in table 4.9.

Where we had information pertaining to their occupation for both parents, we took the ‘greatest’ of these (the lowest in the case of ISCO and Goldthorpe) in any further analyses unless stated, where only

Table 4.8: The Goldthorpe categories

Code	Goldthorpe category	N
1	Higher Controllers	38
2	Lower Controllers	76
3	Routine Non-manual	32
8	Skilled Worker	8
9	Unskilled Worker	14
10	Farm Labourer	1
11	Self-employed Farmer	1
	NA	13
	Total	183

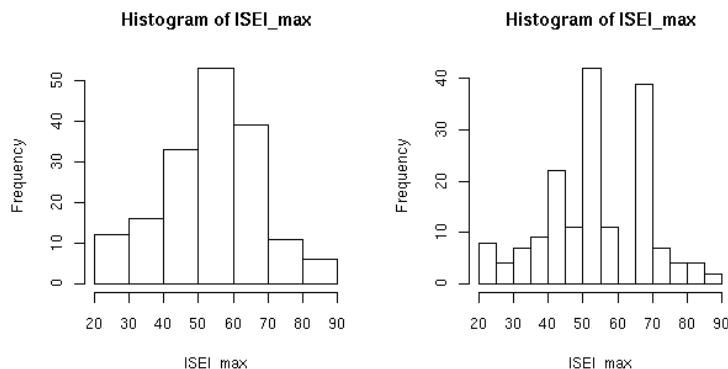


Figure 4.2: The ISEI variable

Table 4.9: ISCO and Goldthorpe classifications

Goldthorpe	ISCO								
	1000	2000	3000	4000	5000	6000	7000	8000	9000
1	15	23	0	0	0	0	0	0	0
2	25	27	24	0	0	0	0	0	0
3	0	0	5	14	13	0	0	0	0
8	0	0	0	0	1	0	7	0	0
9	0	0	0	1	5	0	1	4	3
10	0	0	0	0	0	1	0	0	0
11	1	0	0	0	0	0	0	0	0

the occupation of one parent was given, we used that one. When we use the Goldthorpe classification we shall collapse groups 1 and 2 into the ‘service’ class (N=114), with the remaining groups being put into a ‘non-service’ class (N=56).

A cross tabulation of age against the Goldthorpe (service, non-service) classification with first generation university is shown in table 4.10. The students who are not first generation university are predominantly service class. For those who are first generation, older students figure largely in the non-service class.

There was no significant association of socio-economic status (SES) with age or access route to university. However, there was a significant link of SES with prior attainment (GCSE⁵) (see 4.11) and with first generation at university (4.12), and also a significant link of age with GCSE (4.13).

4.3.3 The sub-samples used in the analysis

As a final point in this section, we detail the number of students in the various combinations across the instruments (4.14). The prescribed assessment for the module (the investigation) inevitably caught a higher proportion of the students; 84 out of the 86 in cohort 04 and 96 out of 97 in cohort 05. For the

⁵Note: “High” in the GCSE measure used in these descriptions refers to at least a “B” grade in both maths and science, low accordingly refers to below “B” in at least one of them.

Table 4.10: Age and Goldthorpe classification

Age	First generation	Non-service	Service
Old	No	1	18
	Yes	21	18
Young	No	5	30
	Yes	27	45

Table 4.11: SES and attainment at GCSE

SES	GCSE	
	high	low
Service	52	54
Non-service	17	35

Table 4.12: SES and first generation university

SES	First generation	
	No	Yes
Service	48	63
Non-service	6	48

other elements of the research, the number of students present at that particular session varied somewhat. This leaves us with an analysis problem. If we only take the students who completed all the elements, the sample size is reduced to 51 for cohort 04 and 70 for cohort 05. On the other hand, we could use as many of the students as we can for any particular analysis (so if we are amalgamating data from pre and post test we would be working with 59 students from cohort 04 and 78 from cohort 05). Either way presents problems. The first reduces the sample, and in a way which may or may not be random. The second uses the largest sample possible but with the risk that different combinations of instruments will exhibit different biases in the absences. We decided to opt for the second course of action, but checked, in the case of major conclusions, that switching to the smallest sample did not materially affect the conclusions we felt able to draw.

4.4 A note about the language of necessity and sufficiency

In a later chapter, and in associated journal articles, we will use the method of data analysis which has come to be known as Qualitative Comparative Analysis (QCA). The basis of this approach is to examine the data with a view to ascertaining whether there are any configurations of potential causes which are either necessary or sufficient for the outcome in questions. We digress briefly here to establish the language used in such an approach.

4.4.1 Sets and logic

A lot of quantitative social research is concerned with correlational thinking, that is, with looking for a measure of association which indicates that two or more variables are linearly related in a certain way. An alternative to this way of thinking is the analysis of necessary and sufficient conditions for a given outcome. Such a way of thinking comes naturally to us. It is often implied in statements such as “anyone with a degree will get a good job”, where a degree might be termed a sufficient condition for getting a good job. Another such statement is “neurosurgeons have good fine motor skills” where it is implied that having good fine motor skills is a necessary condition for becoming a neurosurgeon.

These relationships are not necessarily symmetrical. Sufficiency does not imply necessity and vice versa. In the first, sufficiency-related, statement, it is perfectly possible that other people who have only a lower level of qualification will also get good jobs. A degree may be a sufficient condition, but it does not have to be a necessary one. With regard to the second statement, we can be fairly certain that having good fine motor skills is not a sufficient condition for becoming a neurosurgeon. Other factors

Table 4.13: Age and GCSE

Age	GCSE	
	high	low
old	10	41
young	61	51

Table 4.14: The sub-samples which completed various of the assessments

	Assessment	Cohort 04	Cohort 05
	Registered students	86	103
Q	Biographical questionnaire	86	97
Q+EvPre	Evidence pre-test	76	89
Q+EvPost	Evidence post_test	62	85
Q+Invest	Investigation	84	94
Q+Transfer	Transfer task (Weardale)	66	86
Q+EvPre+EvPost		59	78
Q+EvPre+Invest		76	86
Q+EvPre+Transfer		62	78
Q+EvPre+EvPost+Invest		59	77
Q+EvPre+EvPost+Transfer		51	70
Q+EvPre+EvPost+Invest+Transfer		51	70

will have to be present too, but the fine motor skills are the prerequisite without which it is impossible to work as a neurosurgeon.

Sufficiency and necessity may be conceived of in set-theoretic terms, considering subset relationships. In the first example, we might say that graduates form a subset of the people with good jobs. In the case of sufficiency, then, the condition is a subset of the outcome, or, more precisely, individuals with the condition form a subset of those with the outcome.

In the second example, we might say that those who become neurosurgeons are a subset of all those who have good fine motor skills. Again, we look at a subset relation, but in the case of necessary conditions, the outcome is a subset of the condition.

Sets can be combined through intersection, union and negation. Intersection corresponds to logical AND, union to logical OR, and negation to logical NOT. Such Boolean operators can produce more complex statements which may refer to conditions which are both sufficient and necessary with respect to a given outcome. Going back to our examples, we might state that having a degree OR having rich parents (and nothing else) will lead to a good job, and that having good fine motor skills AND having the appropriate anatomic knowledge are required to become a neurosurgeon ⁶. The following notation is used:

$$\begin{aligned}
 & \text{DEGREE} + \text{RICH PARENTS} \Rightarrow \text{GOOD JOB} \\
 & \text{MOTORSKILLS} * \text{ANATOMY} \Rightarrow \text{NEUROSURGEON}
 \end{aligned}$$

where + indicates logical OR, * indicates logical AND, upper case letters indicate the presence of a factor and lower case letters its absence. Here, then, both equations summarise conditions – or rather, combinations of conditions – which are jointly necessary and sufficient.

⁶We should like to note that all the examples used are of course oversimplified.

Chapter 5

An outline of the teaching programme

As we have noted in earlier chapters, our concern is principally with the understanding and use of evidence in investigative work in science and in our version of scientific literacy. But, of course, we acknowledge the critical role that substantive understanding plays - without it evidence is nothing. The interplay between the two will be the focus of the research into students' ability to carry out practical investigations. In this chapter we shall describe the teaching programme in some detail. The School of Education at Durham has a number of courses aimed at Initial Teacher Education. One of these is the undergraduate BA (Education). It is a 3 year course for intending primary teachers.

5.1 The BAEd course at Durham

The 3 year BAEd course for primary teachers aims to develop students' own understanding in the core subjects of the primary curriculum as well their understanding of educational and pedagogical issues. Students take the equivalent of 6 modules each year in addition to their practice in schools. The science element consists of a module each year that focuses on students' own knowledge and understanding of key ideas in science. In the 1st year, students follow a module that addresses key ideas in chemistry and physics. In the 2nd year the module is in two distinct parts: teaching about procedural understanding and the key ideas of biology. This paper describes the part of the module that aims to teach for procedural understanding and is henceforth referred to as the evidence module. In the third year, there are further science modules which are very focused on the practicalities of teaching in primary classrooms.

5.2 The modules targeted at substantive understanding

One and a half modules are aimed at physics, chemistry and biology, each of them taking roughly half of the teaching time of one module or some 15 hours contact and considerable more in directed time. The physics element, which is relevant to the investigations we use here, covers a small number of what we believe to be the key ideas. In particular it devotes almost a half of the time available to force and motion. The approach is to start absolutely from scratch as, in our experience, the brief time allocated to the study of this area in the UK national curriculum results in nothing more than confusion and misunderstanding. The reader will get some idea of the approach we adopt by referring to the relevant appendix where we present some examination questions.

5.3 The Year 2 evidence module

We have taught students about procedural understanding for many years based on research we have carried out in Durham and materials that we have developed. The syllabus described, which now constitutes part of the 2nd year module, while mainly taught by one of us (RR), must be credited to many colleagues over the years.

The part of the module that aims to develop a student's procedural understanding lasts 9 weeks. Each of the 9 teaching weeks has the format of a 1 hour lecture followed, for the first seven weeks, by 1.5 hours

of workshop activities in classes of about 25. The workshops in the final 2 weeks are allocated to the assessed practical investigation and do not involve any teaching. So each student has 19.5 hours of direct face-to face teaching in total. In addition to the recommended course text (Gott and Duggan [2003a]), materials are provided on-line to support students' independent study. These materials include copies of the lecture presentations, short self-tests and activities (similar to materials found in the Collins resources (Gott et al. [1997, 1998, 1999b]) and an interactive IT-based investigation (which was the forerunner to Folens' Building Success in Sc1, Gott and Duggan [2003b]).

5.4 The rationale behind the teaching

The students are training to become primary teachers. We want the students to understand evidence so that they can later make decisions about teaching science as well as developing their own understanding of science. We have argued that this requires the students to understand the concepts of evidence. But how to teach so that students can really get to the bottom of things so that they understand evidence is not an easy question. We could have taught them through them conducting lots of investigations, or we could have set the ideas in the context of contemporary issues that they analysed, or a combination of both. An alternative approach would be explicitly to teach the concepts of evidence. We decided to try the latter. It is an empirical question as to whether this is the best approach but with limited teaching time we felt that this could be efficient. The resultant evidence module is described in more detail in the sections below.

5.4.1 The logical sequence

We have distinguished the ideas that are the learning outcomes of our teaching, the knowledge-base of evidence, from the way in which those ideas can be applied to solve practical problems in an investigation or to engage with issues by asking questions about the quality of the evidence behind a public claim. This first stage of our research took a somewhat extreme stance in that we deliberately took the 'concepts of evidence' as a *syllabus* in isolation from any extended investigations or authentic issue of more public relevance. Each idea was exemplified using short tasks that were chosen deliberately to illustrate the idea in as clear and uncluttered way as possible, rather than for any real or apparent authenticity. There was, therefore, an attempt to make the teaching contexts relevant to the idea being taught, rather than to the students' personal experience or current debates. The argument here is that, if this were to prove effective, then we would have a parsimonious teaching scheme. If it did not, then we would modify the teaching on the spot if ethical issues were involved (but this was not needed), or develop teaching patterns for subsequent cohorts by, for instance, using more long investigations, or by teaching with current issues, as dictated by the findings of the current work.

We hypothesise that:

- carrying out a good investigation will depend on an understanding of the concepts of evidence (as well as an understanding of the substantive ideas of science), organised through an understanding of the design of investigative tasks (looking forward)
- and that being able to ask forensic questions about the evidence behind a public claim in a socio-scientific issue will depend on an understanding of the concepts of evidence and of design issues (looking back), figure 5.1.

In the evidence module students will be taught the concepts of evidence and will be assessed, at the end of teaching, on their ability to conduct a lab-based investigation. They will then be presented with the Issue task where they are asked to question the scientists' claim in a local socio-scientific issue.

5.5 The evidence module

The evidence module is, as far as we know, a novel way of teaching for a procedural understanding. We describe it in some detail here so that readers can distinguish our teaching from the many other

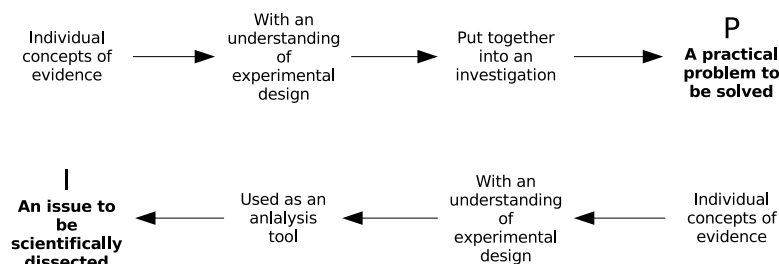


Figure 5.1: The ideas behind the teaching sequence

approaches, such as ‘process and skills’ approaches, inquiry and argumentation which have a different curriculum emphasis.

Students’ procedural understanding is developed by explicitly teaching them the concepts of evidence required to conduct a lab-based investigation. Since the module aims to develop these ideas the examples used as contexts for teaching and practical work are deliberately very simple so that the substantive ideas required to solve problems are kept to a minimum so as not to distract from the procedural ideas being taught. They are similar to contexts that could be used in primary science teaching and therefore are of relevance to the students.

The ideas are the focus and ‘learning outcomes’ of the lectures and are developed further with the students in the workshops. In the workshops activities include:

- Exploratory practicals to familiarise students with the procedural context,
- Practical to illustrate specific procedural ideas,
- Parts of investigations where students can apply some of the ideas they have learned,
- A whole investigation, as a summative assessment,
- Whole class and small group discussion focused on the procedural ideas,
- Video
- Structured worksheet activities mainly in the form of question sheets, based on activities to be found in the Collins ‘Science Investigations Packs’(Gott et al. [1997, 1998, 1999b]).

This part of the module is assessed by the students planning and conducting an investigation: For the 04 cohort, this was ‘Does the material a ‘helicopter’ is made from make any difference to the drop time?’; for the 05 cohort, the investigation was ‘How much do different surfaces affect how easy it is for a shoe to slide?’. All the teaching therefore focuses on the procedural ideas within the context of ‘looking forward’; the lectures and workshop activities concentrate on the ideas the students will need to understand so that they can conduct their own open-ended investigations in the context of a lab-based investigation. An approximate sequence of events is described in table 5.1.

This teaching sequence was the same for both cohorts with just one difference. The 05 cohort had an open-ended investigation in the Week 1 workshop. The ideas listed in the table for Weeks 1 & 2 were addressed in Week 2 for the 05 cohort. By way of further illustration, we shall briefly describe two sessions so that the reader has some idea of what the course comprises.

Table 5.1: The course: a brief overview

Week	Lecture 1 hour	Workshop 1.5 hours
1	Introduction to the module. Introduction to Procedural Understanding. Role of practical work in science education.	Exploratory practical in small groups to introduce basic ideas of designing an investigation. Students are provided with 'beams' of three different materials (perspex, wood and MDF) of the same length but different width and thickness, blocks of different dimensions to act as pillars, a selection of masses and length-measuring equipment. In groups they investigate how the material affects the sag of a beam bridge. The subsequent discussion introduces students to validity of design and some measurement issues. The decisions students have made is explicitly linked to the knowledge-base of the concepts of evidence.
2	Presenting data: tables and graphs. Introduction to Science Investigation Workshop on intranet.	Beam bridge practicals in small groups to extend design issues: relationship between the independent variable (IV), the dependent variable (DV) and control variables (CVs); range and interval of the independent variable. Planning 'spring board man' investigations for following week: how does the mass applied to a 'spring board' affect the height that a 'man' jumps?
3	Validity of design. Values of Control Variables.	Practical in small groups: 'spring board man' investigation focusing on the range of independent variable, values of control variables and how they affect the dependent variable; repeated readings; making a measuring instrument.
4	Measuring instruments. Uncertainty.	Video about measuring blood pressure to introduce students to relationships underlying measuring instruments, factors affecting the use of an instrument, accuracy and precision and other terminology. Activity: examining measuring instruments - students are presented with a range of measuring instruments and consider factors that would affect the reliability of measurements.
5	Variation in repeated readings. Anomalies? Frequency histograms. Normal distribution.	Practical in small groups: to design a way to measure the absorbency of a paper towel and to do repeated readings. Discussion about repeated readings; the causes of variation and how variation can be minimised.
6	Descriptive statistics: Standard Deviation and Standard Deviation of the Mean (Standard Error). Working iteratively/responding to data.	Activities to develop understanding of descriptive statistics. Use of IT for data presentation and data analysis.
7	Using descriptive statistics to draw conclusions. Working iteratively/responding to data.	Practical in small groups: putting it all together in a whole investigation 'How does the surface onto which a drop is released affect the size of the splash?' Class discussion at intervals during investigation to consider application of ideas and how they inform the next stage of the investigation.
8	Revision of key ideas in preparation for Assessed practical. Explaining 'the thinking behind the doing'.	Lab open for 3 hours: individual assessed investigations using categoric independent variable.
9	The implications of the concepts of evidence for teaching primary science.	Lab open for 3 hours: individual assessed investigations using categoric independent variable.

5.5.1 Examples of teaching intervention

Table 5.2: The course: Session 3

Week	Lecture 1 hour	Workshop 1.5 hours
3	Validity of design. Values of Control Variables.	Practical in small groups: ‘spring board man’ investigation focusing on range of Independent Variable, values of Control Variables and how they affect the Dependent Variable. Repeated readings. Making a measuring instrument.

Week 3 (table 5.2). In the week prior to the 1.5 hour workshop the students, working in groups of 3s, had been presented with the problem:

‘How does the mass applied to a spring board affect the height the man jumps to?’

They were shown the basic arrangement of apparatus. A metre stick could be clamped to the bench with some overhang to create a ‘spring board’. A ‘Lego’ man was put on the springboard. A mass could be suspended from the spring board with string. When the string was cut, releasing the mass, the height that the ‘Lego’ man ‘jumped’ could be recorded. Students were provided with a range of apparatus:

- Various metre sticks made from metal, wood and plastic with different resolutions of scale on them i.e. millimetre, centimetre and decimetre for measuring and to create a ‘spring board’,
- Tape measures, 12 inch rulers,
- Assortment of masses: 10g slotted masses on hangers, 100g slotted masses on hangers, 0.5kg masses, 1kg masses,
- G-Clamps, Boss and Clamps,
- Assorted ‘Lego’ men: various sizes and shapes,
- Selection of strings: thin, thick, nylon etc.,
- Scissors,
- Sellotape and blutack,
- Top pan balances.

In advance, students were asked to identify the independent (IV: mass applied) and dependent variables (DV: height) and any control variables (CVs: material of spring board, flexibility of spring board, amount of overhang, position of man, position of hanging masses, etc.). In the workshop they worked in groups to gather data to answer the problem. At intervals the class was brought together by RR to discuss the decisions they were making and to remind them to record data in tables, plot graphs as the data was collected, note the values of CVs that they decided on. Each group made different decisions; the important thing was to be able to recognise how they had to make decisions in response to the data as it was collected:

- The importance of early ‘trials’ was to get a feel for the problem.
- What other Control Variables had been identified once they started to run trials? How would they control their values?
- What range of values of the Independent Variable could be used? The importance of considering this in relation to the maximum measurable height of the Dependent Variable as well as the values set for the Control Variables was discussed. Students were encouraged to see what other groups had done to solve the problem. They realised there were many valid ways to answer the question and that the data being collected depended on those decisions.

- What interval of values of the Independent Variable was necessary to answer the question? The importance of an emerging relationship on a line graph was emphasised.

Students soon realised that repeated readings gave different values of the Dependent Variable. They were encouraged to repeat their readings until they felt they had enough to ‘capture’ any variation they might find and record the repeats in a table. All the repeated readings were plotted as points on their graphs. Students were encouraged to discuss how best to summarise this variation for each value of the Independent Variable and produce a graph that best showed the relationship between the Independent Variable and the Dependent Variable. Once students had produced a line graph they were given a block of unknown mass and asked to find its mass using the spring board ‘as a measuring instrument’. In groups they discussed the task and carried it out. They then checked the mass on top pan balances around the lab and considered the uncertainty inherent in the spring board as a measuring instrument. Follow up work involved reading the chapter about Measuring Instruments in the recommended course text Gott and Duggan [2003a].

Table 5.3: The course: session 5

Week	Lecture 1 hour	Workshop 1.5 hours
5	Variation in repeated readings. Anomalies? Frequency histograms. Normal distribution.	Practical in small group: to design a way to measure the absorbency of a paper towel and to do repeated readings. Discussion about repeated readings.

Week 5 (Table 5.3). This workshop involved students working in small groups to make repeated measurements. They were told they had to devise a way of measuring the absorbency of paper towels. The students had to design a measurement method which they could refine through trial and error. They did not have to conduct a whole investigation into factors that might affect absorbency. The purpose was to collect repeated readings until they felt they had collected enough. Students were provided with:

- Kitchen roll,
- Scissors,
- Rulers with different scales,
- Assorted containers: trays, washing up bowls, plastic containers,
- Assorted volume measurers: Beakers, measuring jugs, measuring cylinders, conical flasks, volumetric flasks, pipettes,
- Assorted scales: kitchen scales, top pan balances with different resolutions, balances.
- The validity and reliability of the many designs was discussed with each group and their designs were often refined following the discussion.

Following data collection the class discussed with RR why their data showed variation. Measurement uncertainty was discussed and consideration was given to how it could be reduced. Anomalous data were considered - were they indicative of greater natural variation than students had collected in their sample of repeated readings or could anomalies be explained in some way? A discussion followed as to how the anomalous data might then be treated. Students were asked to summarise their varied data and consider the limitations of using the mean, mode or median and the range.

5.6 A ‘task analysis’ using observations from teaching

5.6.1 Introduction

The tasks given to the students and used in this research are detailed elsewhere. In summary they involved:

- a written assessment of the concepts of evidence - the Written evidence test,
- an open-ended investigation where the concepts of evidence could be applied in a practical problem solving context - the Investigation,
- the application of the concepts of evidence to a local socio-scientific issue - the Issue task.

In an earlier chapter we argued for the relationship between concepts of evidence, investigations and scientific literacy; that the concepts of evidence are the knowledge-base used when solving practical problems in science and when looking back on others' claims to evaluate them. The tasks we have given the students attempt to measure each of these aspects of the relationship. Other research, detailed elsewhere and our own experience during this research suggests that students are differentially successful on these tasks. We have, therefore, attempted to look at other variables that might affect their success.

These are some insights gained from the experience of teaching and administering the three student tasks used in this research. We present this descriptive account to provide some contextual information for our selection of variables that might have differentially affected students' performance on the different tasks.

5.6.2 The Written evidence test

The written evidence test has been used as a pre-test and a post-test in this research. As a pre-test given to students just before teaching commences it is likely to appear to the students as a novel test. Although (all too?) familiar with written tests in science from their experience of GCSE and A level, it is unlikely that the students will have encountered questions that exclusively target their understanding of ideas about evidence in the form used here. Students were advised that the pre-test was being given to them for research purposes and would not form part of the assessment of the module but that their answers would be used to enable the tutor (RR) to better understand the cohort. Most students were engaged with the test, many taking time to read the questions carefully. Some students completed the pre-test quickly. Several students commented, on completion, that they had never experienced such questions before. Some were very uncertain about how well they had done; their lack of confidence in science was evident.

After the students had been taught the module, the written evidence post-test used in this research, would, by then, be in a relatively familiar format to the students. The questions target ideas that the students had become aware of during the module and were set in simple contexts where any requirement for understanding of any substantive ideas is relatively low. The students perceived the questions as being relatively straight forward after the module and seemed to take the post-test in their stride. The post-test was not part of their assessed work for the module and they did not feel threatened by it. A few commented that they feared that they have forgotten the ideas since they were taught them 2-3 months earlier - they were not told of the post-test in advance and had no opportunity to revise.

The questions in the test are relatively tightly targeted: students can focus on the few ideas required to answer each question. Each question is relatively short so this limits the number of ideas the student will have to hold in his/her head to be able to frame their answer. The test clearly targets lower levels of the Bloomian taxonomy: knowledge, comprehension and application in tightly defined contexts. Perhaps as important is the convergent nature of the questions. This fits with students' experience of schools in the UK where the cramming of factual knowledge for short questions in the ubiquitous attainment tests, much loved of Government, and externally set examinations is the norm.

5.6.3 The investigation

The investigation provides students with the opportunity to apply their understanding of the Concepts of Evidence in a novel problem-solving context similar to activities they might come across in primary science classrooms. For the 04 cohort the investigation was administered only after teaching the module. The 05 cohort did an investigation both before and after teaching. The specific context of the investigation was not used in the teaching.

Although scaffolded practical activities involving parts of investigations are used to teach ideas in the module, the 'investigation' is the first time they have individually put all the ideas together to design

the investigation, evaluate and modify their approach as they collect data and interpret their results and write their account. Whilst such tasks are theoretically within the National Curriculum for science, they have become so routinised in practice (Roberts and Gott [2003]) that these open-ended tasks are a genuine novelty to all our students, which they often comment on.

The investigation involves students applying their understanding of the concepts of evidence in a novel open-ended investigation. To be able to conduct the investigation successfully they have to select the relevant concepts of evidence for each stage of the task, synthesise them to address the problem and evaluate the quality of the evidence they collect. Thus they have to hold together ideas and decisions from the various stages of the investigation to attempt to solve the problem. The investigation targets higher levels of the Bloomian taxonomy than the written evidence test. Other factors involved in the investigation and which distinguish it from the written task include the openness, and therefore relative complexity, of the question; the high stakes nature of the summative investigations (it contributes 25% of their marks for science for the year) where students were asked to work individually on the task; the students' use of practical apparatus and their ability to be organised in the lab and their confidence in the use of IT.

The investigation, whether presented to the students before or after teaching, is always seen as threatening by the majority of students. For many students it is seen as an extremely stressful experience. It is the topic of conversation and concern from the time it is announced until the word processed write up is complete. Students approach the task with various degrees of interest. The summative nature of the investigation at the end of the teaching means that most try hard but for some the task itself is treated in a fairly perfunctory fashion and repeated readings are seen as decidedly tedious. Others get really engaged, seemingly over and above the requirements for a good mark, and seem to relish identifying and solving issues that arise. For many students the openness of the task is daunting. Teaching staff do not answer student questions related to the investigation during the task. Although the students work individually to design their own investigation and collect data the majority of students seek support from each other at some stage during the investigation.

Students have commented that this is the most 'thinking' they have ever had to do while at university! Unlike more traditional essay writing by which they are more frequently assessed - where students can pull together others' writing and ideas to structure their argument - the students feel very 'exposed'. They feel it is very much up to them as individuals to solve the problem and there is nothing else (like other authors' works) that they can fall back on. Many comment that they have never had to work like this before in science in school - things have always been 'closed down' for them (by them having done/seen a similar practical before; by having very few choices to make e.g. there was no choice of what measuring equipment to use since only one sort was put out for them; by them having 'writing frames' or prompt sheets to structure each stage of the process).

Students tend to cope with this open-ended task in various ways. A few students every year are seemingly, almost completely, overwhelmed by the task. In discussion with them they report that they find science modules difficult; their self-efficacy is very low. They seem unable to close down all the possibilities open to them to attempt to answer the question and collect masses of unstructured data that does not lead to an answer. If you talk to them they might know individual ideas about evidence but putting them together in such an unstructured task seems totally beyond them. Their approach we have typified as 'divergent'.

Other students regain control by defining the problem in advance of coming to the lab and often have designed a detailed method they are going use to find a solution (sometimes with reference to other sources of similar practicals such as science texts or the web). Their approach to the openness of the task is to close it down before starting into something with which they are more familiar - almost a 'recipe' practical and often with a desire for 'the right answer'. These students tend to work determinedly, often without much discussion with others, and leave the lab relatively quickly. Hardly any issues are 'problematised' and they tend to apply the ideas about evidence very routinely. We have described their work as 'linear'.

A third approach is taken by students who work more iteratively. They tend to 'play' with the equipment and collect data from small trials to better understand the problem and then manage to come to a decision about how to proceed in a more focused way to answer the problem that they have sensibly and justifiably closed down. They engage with the data they have collected, reflect on it to decide on its quality and make further decisions about what to do based on their evidence. Some students do

this more routinely in parts than others - they know, for instance, that good data has a low Standard Deviation of the Mean (Standard Error) and work to reduce it even if their evidence already enables them to answer the question. We have described this overall approach as 'iterative'.

In addition to the students feeling that they have never had to do so much thinking before, the practical organisation and skills needed to do the investigation are daunting for some. Despite having been taught about using tables as organisers for their data, results are often jotted on disorganised bits of paper. Others find manipulating the apparatus a very time consuming and frustrating task. For instance, using a boss and clamp on a retort stand to set something at a fixed height and away from obstructions proves very frustrating for some.

The other factor that causes great unease among some students is their confidence and abilities using spreadsheets; the incorporation of data and graphs from the spreadsheet in their word processed account adds to their concerns over the task. For some, file management on the computer as they work through various stages of the task is obviously causing problems. There are always a few accounts where students have been unable to plot the data as they wanted it: axes are incorrectly spaced or labelled; descriptive statistics not required in the account are included because they are unable to edit the output etc.. Students are advised not to let the IT prevent them from completing the task and that hand drawn graphics are acceptable. Most persisted with the IT, perhaps because it is a skill they will have to master before qualifying as a teacher.

The final summative percentage mark gained by the students for the investigation sometimes causes some consternation. Students do not always get the sort of mark they would expect based on their more usual essay-based assessments. Every year the investigation mark throws up some surprises; students who do well on other assessments who get very much lower marks on the investigation and some who get the highest mark they have ever achieved on the investigation.

5.6.4 The Issue task

The 'Issue task' is another context entirely. The intention is to offer an everyday 'issue' to the students to see what sorts of reaction it provoked before and after the teaching. We hope to see application of the Concepts of Evidence in this quite different setting, 'looking back' on the claims the experts had made. It was conducted prior to teaching with the 05 cohort, while they were in Year 1 of the degree programme, and some 3-4 months after teaching with both the 04 and 05 cohorts - when they completed the task they had not revised or specifically reminded themselves of the Concepts of Evidence prior to the task.

It is worth noting again that during the module the students have been taught some of the concepts of evidence explicitly. Due to the limited teaching time available and the focus on students conducting their own lab-based investigation, the teaching focused on those Concepts of Evidence most emphasised in lab-based manipulative contexts, and all the examples used during the teaching were lab-based. Thus the Issue task, in a context where variables could not be manipulated to ensure the collection of valid data, would appear very different to anything discussed in class. Some students may have had prior experience of fieldwork at school.

The task involved 'looking back' on others' research. The task was presented to the students as 'local' - many would have been aware to some extent of the Pennine dales and a few could remember the issue itself, but many would not have had first hand experience of the rural context and might therefore have had difficulty putting themselves in the residents' position (as parents, home owners, farmers etc.). Similarly they would have had a range of interests in such issues - some may have been interested in issues of pollution and the concerns of the residents and might have been quite 'green', while others may have had little interest in such matters. They were simply asked to write down their initial thoughts when they were told about the issue and then were prompted to write down questions they would like to ask the scientists involved. Their ability and willingness to raise questions may have been affected by their general attitude to 'official' pronouncements: some may have been naturally sceptical while others may have been more accepting of 'authority' or expertise. Some students may already have had experience of writing to planning departments etc. about developments in their own locality while others may not have realised that this consultation is part of the local democratic process.

All the students asked questions when they were prompted. The nature of the questions varied: questions differed in the nature of the focus (on evidence, on substantive issues, on economic issues) as

well as in how specific they were. Many students wrote questions that were not evaluating the problem (such as whether the cement works would provide further updates to the locals on any future change of fuel) or actually just rephrased a question previously asked - perhaps an indication of trying to complete the task but actually having difficulty thinking of many things to ask. Comments from the students suggested that they found the task quite interesting, particularly when administered after the module. The task was seen as non-threatening and students engaged with it. The format was unlike that of any formal assessment; students were given an open activity where they were asked their views and asked to raise questions.

5.6.5 What issues does this raise for the research and its explanatory framework?

The concepts of evidence are the knowledge-base required for a procedural understanding and are the ideas specifically taught in the module, using both practical and non-practical tasks. These ideas can be assessed by written evidence tests. It is hypothesised that explicit teaching of the ideas results in students understanding them more so that they can recall, comprehend and apply the ideas about evidence in the context of the test. Students who made big gains between pre- and post-test could reasonably be expected to be students with other good attainment measures - they learn things well. But since the students did not revise for the post-test which occurred several months after teaching, perhaps it rewards students with rather deeper learning than the rote recall on traditional assessments.

Research at a local school indicates that 'quick-witted' pupils performed well on the evidence test and that it 'tapped into' some abilities that were somewhat distinct from usual school assessments. It is hypothesised that performance on the evidence pre-test is less well correlated with social class than traditional assessments (GCSEs, A levels, 'access', and - probably - year 1 exam results).

The concepts of evidence are ideas used by pure and applied scientists (Gott et al. [1999a], Roberts and Gott [2003]) in their practical problem solving work. It is hypothesised that students with a better understanding of concepts of evidence will be able to use these ideas in the context of an investigation. We will be able to compare this with the pre- and post- sliding shoes investigations with 05 cohort and look for correlations between the post-test and the summative investigation for both cohorts. Practical problem solving by carrying out an iterative investigation may also be influenced by students' ability to deal with complexity and uncertainty. It could be argued that older students with wider real-life experience may have more experience of this compared with students straight from the school system that rewards 'correct' answers in relatively closed contexts.

Tytler et al's (Tytler et al. [2001a,b]) research has shown that an understanding of concepts of evidence enables the public to ask questions of others' evidence in a real environmental context. It is hypothesised that the number of questions about the evidence raised will increase after teaching and, more particularly, that the number of specific 'Concepts of evidence' coded questions will increase. The students' willingness to unquestioningly accept the experts' findings could be argued to be inversely related to their scepticism and ability to deal with complexity and uncertainty, so this too may be age related.

Since an understanding of concepts of evidence (post-test) is arguably a necessary pre-requisite for performance on both the investigation and the Issue task, and since it could be argued that both performances require the student to ask questions using concepts of evidence (of their own data as they solve the investigation or of others' data) it could be hypothesised that results from the Issue task and investigation are correlated. However, the investigation was part of the students' summative assessment and was therefore treated very diligently by the students, so their total percentage mark may be more a reflection of their diligence. Holding the concepts of evidence in their heads and using them to ask questions as they did their investigation would result in investigations being coded as 'Iterative' - thus their approach to the investigation might be correlated with their ability to deal with complexity and uncertainty as well as their understanding of concepts of evidence in the post-test. 'Linear' approaches that close the question down to a task more akin to previous school experience and 'Divergent' approaches to the investigation could be seen as indicators of being unable to deal with open-ended tasks and possibly the associated complexity and uncertainty. These may be associated with age.

Both the Issue task and the investigation are examples of performance in novel contexts where learning needed to be applied differently from the way in which it was taught.

5.6.6 Some tentative hypotheses derived from teaching experience

From our experience of teaching this course over a number of years, illustrated briefly in the above sections, we can make some tentative suggestions, which are in no particular order, as to what factors might be important for those students who are successful at one or more of the tasks.

1. Disposition. Determination and willingness to engage with novel ideas c.f. spoon feeding.
2. A wish to have neat convergent answers - a 'security blanket' - c.f. those who are willing to accept uncertainty. Worries about new ideas, particularly new ideas that are about uncertainty which cut directly across their experience of science in schools as providing the 'right' answer. These ideas do not sit comfortably with the popular perception that science is absolute, and scientists are only unsure due to incompetence or lack of knowledge rather than the inherent uncertainty always associated with measurement.
3. Test-wisdom. Many students come from a heavily crammed course and accept surface knowledge as the norm.
4. Ability - in maths and science.
5. A disposition to be critical of authority statements.
6. Prior attainment in maths and science.
7. Background factors such as first generation university and social class which, through notions of cultural capital, bring a certain attitude of mind and experience of academic learning.
8. Ideas about evidence (procedural understanding in our terms) are less context dependent than substantive ideas and, moreover, might act as bridging ideas (scaffolding) making transfer of substantive ideas across contexts more likely.

Some of these are unmeasurable. Others have not been measured due to lack of an appropriate measuring instrument. Most commonly, however, the length of the module and the busy nature of the course as a whole precluded any additional time for data collection. Given what we have been able to measure, we can identify a number of research questions which can either be addressed directly, or on which some light can be shed indirectly and these are dealt with in the next section.

5.7 Research questions

Some of these have been answered in small scale exploratory case studies. Others are ongoing.

In no particular order:

1. Does students' understanding of the Concepts of Evidence improve after explicit teaching?
2. Can students who are explicitly taught the concepts of evidence so that they can construct a claim in an open-ended, lab-based investigation ask qualitatively different questions about a science claim in a local socio-scientific issue after teaching compared with before?
3. Is an understanding of substantive ideas a necessary or sufficient condition for good performance on an open-ended investigation?
4. Is an understanding of procedural ideas a necessary or sufficient condition for good performance on an open-ended investigation?

If an understanding of procedural ideas is important:

5. Which elements of procedural understanding benefit most/least from the short targeted teaching programme, which explicitly teaches ideas about evidence?
6. Does an improved understanding of procedural ideas enable students to adopt a more sophisticated approach to data collection?

Appendix A

The evidence test

The test consisted of 17 items, each of which was broken down into a number of parts. Some items were multiple choice, others were short free response in style. The items were not, therefore, of either equal length or equal difficulty. A decision was taken, as a consequence, to weight the individual parts of each item, and then to weight each item for addition into the item and test totals respectively.

Each item was allocated a maximum mark of 10. The parts of the item were then allocated marks, subjectively, corresponding to our judgement as to their level of difficulty. This resulted in 17 item scores, each with a maximum mark of 10. Then each item was allocated a weighting ranging from 1 to 4. The '1' items were those requiring a single answer, and were relatively easy. The '4' items, on the other hand, were more complex, took longer to answer and attacked more difficult ideas.

The item weightings are shown in table A.1. The third column indicates the questions which were used in the post-test.

A.1 Coding of the test

A.2 Tests for normality and item statistics

A.2.1 Normality

The pre-test is acceptably normal with a p-value bordering on 0.05. We shall take this as being sufficient for further analysis.

Pearson chi-square normality test

data: EvPre

$P = 15.5789$, p-value = 0.07621

`lillie.test(EvPre)`

Lilliefors (Kolmogorov-Smirnov) normality test

data: EvPre

$D = 0.1027$, p-value = 0.04592

A.2.2 Various statistics related to the evidence pre and post-tests

1. Table A.2: The evidence pre-test - item correlation matrices.
2. Table A.3: The evidence post-test - item correlation matrices.
3. Table A.4: The pretest: item statistics.
4. Table A.5: The pretest: summary statistics.
5. Table A.6: The post-test items within the pre-test: item statistics
6. Table A.7: The post-test items within the pre-test: summary statistics

Table A.1: The item weightings in the test score

Item	Weighting - pre	Max weighted marks per item	Weighting - post	Max weighted marks
1	3	30		
2	3	30		
3	1	10		
4	2	20	2	20
5	4	40	4	40
6	4	40		
7	3	30		
8	1	10	1	10
9	1	10	1	10
10	2	20		
11	3	30	3	30
12	2	20		
13	1	10		
14	2	20		
15	1	10		
16	2	20	2	20
17	4	40	4	40
Totals	39	390	17	170

7. Table A.8: The post-test: item statistics

8. Table A.9: The post-test: summary statistics

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17
T1	1.00	-0.03	0.13	0.24	0.24	0.20	0.07	0.09	0.10	0.17	0.21	0.11	0.01	0.30	0.13	0.14	0.20
T2	-0.03	1.00	0.09	0.10	-0.01	0.11	0.02	0.06	0.11	0.02	0.07	-0.10	0.07	0.07	-0.04	0.05	0.11
T3	0.13	0.09	1.00	0.01	0.02	0.07	0.03	0.16	0.14	0.11	0.05	0.11	0.07	0.05	0.07	-0.08	0.13
T4	0.24	0.10	0.01	1.00	0.30	0.25	0.14	0.17	-0.07	0.04	0.03	0.05	0.11	0.33	0.21	0.15	0.29
T5	0.24	-0.01	0.02	0.30	1.00	0.12	0.11	0.18	0.05	0.18	0.11	0.11	0.09	0.23	0.23	0.13	0.24
T6	0.20	0.11	0.07	0.25	0.12	1.00	0.11	0.18	0.10	0.05	0.12	0.09	0.07	0.15	-0.12	-0.01	0.15
T7	0.07	0.02	0.03	0.14	0.11	0.11	1.00	-0.08	0.02	0.00	0.05	-0.02	0.06	-0.02	-0.03	0.03	0.07
T8	0.09	0.06	0.16	0.17	0.18	0.18	-0.08	1.00	0.21	0.08	0.05	0.01	0.14	0.25	0.10	0.19	0.13
T9	0.10	0.11	0.14	-0.07	0.05	0.10	0.02	0.21	1.00	0.17	0.28	0.08	0.05	0.07	0.03	0.15	0.03
T10	0.17	0.02	0.11	0.04	0.18	0.05	0.00	0.08	0.17	1.00	0.20	-0.03	0.02	0.01	0.22	0.14	0.12
T11	0.21	0.07	0.05	0.03	0.11	0.12	0.05	0.05	0.28	0.20	1.00	0.04	0.06	0.01	0.01	0.14	0.15
T12	0.11	-0.10	0.11	0.05	0.11	0.09	-0.02	0.01	0.08	-0.03	0.04	1.00	0.08	0.18	0.11	0.05	0.12
T13	0.01	0.07	0.07	0.11	0.09	0.07	0.06	0.14	0.05	0.02	0.06	0.08	1.00	0.01	0.09	0.11	0.02
T14	0.30	0.07	0.05	0.33	0.23	0.15	-0.02	0.25	0.07	0.01	0.01	0.18	0.01	1.00	0.20	0.22	0.36
T15	0.13	-0.04	0.07	0.21	0.23	-0.12	-0.03	0.10	0.03	0.22	0.01	0.11	0.09	0.20	1.00	0.23	0.11
T16	0.14	0.05	-0.08	0.15	0.13	-0.01	0.03	0.19	0.15	0.14	0.14	0.05	0.11	0.22	0.23	1.00	0.15
T17	0.20	0.11	0.13	0.29	0.24	0.15	0.07	0.13	0.03	0.12	0.15	0.12	0.02	0.36	0.11	0.15	1.00

Table A.2: Evidence pre-test: correlation matrix (N=137)

Table A.3: Evidence post-test: correlation matrix (N=137)

	T4	T5	T8	T9	T11	T16	T17
T4	1.00	0.35	0.17	0.11	0.17	0.20	0.11
T5	0.35	1.00	0.15	0.05	0.15	0.07	0.14
T8	0.17	0.15	1.00	0.07	-0.06	-0.01	0.02
T9	0.11	0.05	0.07	1.00	0.03	0.00	0.06
T11	0.17	0.15	-0.06	0.03	1.00	0.10	0.09
T16	0.20	0.07	-0.01	0.00	0.10	1.00	0.20
T17	0.11	0.14	0.02	0.06	0.09	0.20	1.00

Table A.4: Evidence pre-test: item statistics (the Cronbach function in R)

ITEM.STATISTICS	Item.Mean	Item.Var	Item.SD	Item.Tot.Cor	Item.Rem.Alpha	N.Cases
T1	5.8	5.9	2.4	0.366	0.62	165
T2	7.2	5.3	2.3	0.122	0.65	165
T3	4.3	11.4	3.4	0.182	0.64	165
T4	6.3	10.2	3.2	0.374	0.61	165
T5	8.0	10.7	3.3	0.370	0.61	165
T6	7.5	4.0	2.0	0.286	0.63	165
T7	3.8	13.0	3.6	0.089	0.66	165
T8	6.0	11.2	3.3	0.302	0.62	165
T9	4.0	6.4	2.5	0.224	0.63	165
T10	5.9	10.2	3.2	0.214	0.64	165
T11	2.5	3.0	1.7	0.241	0.63	165
T12	5.9	3.0	1.7	0.147	0.64	165
T13	6.6	7.1	2.7	0.156	0.64	165
T14	4.2	8.2	2.9	0.374	0.61	165
T15	6.3	1.1	1.1	0.252	0.64	165
T16	4.7	3.4	1.8	0.267	0.63	165
T17	5.1	9.2	3.0	0.385	0.61	165

Table A.5: Evidence pre-test: summary statistics (N=165)

TOTAL.STATISTICS	Cronbachs.Alpha	Standardized.Alpha	Total.Mean	Total.Var	Total.SD	N.Items
TOTAL	0.64	0.66	94	314	18	17

Table A.6: Evidence pre-test: statistics for the post-test items (the Cronbach function in R)

ITEM.STATISTICS	Item.Mean	Item.Var	Item.SD	Item.Tot.Cor	Item.Rem.Alpha	N.Cases
T4	6.3	10.2	3.2	0.30	0.49	165
T5	8.0	10.7	3.3	0.34	0.47	165
T8	6.0	11.2	3.3	0.29	0.49	165
T9	4.0	6.4	2.5	0.16	0.54	165
T11	2.5	3.0	1.7	0.21	0.52	165
T16	4.7	3.4	1.8	0.28	0.50	165
T17	5.1	9.2	3.0	0.32	0.48	165

Table A.7: Evidence post-test: summary statistics for the post-test items (N=165)

TOTAL.STATISTICS	Cronbachs.Alpha	Standardized.Alpha	Total.Mean	Total.Var	Total.SD	N.Items
TOTAL	0.54	0.54	37	100	10	7

Table A.8: Evidence post-test: item statistics (N=137)

	Item.Mean	Item.Var	Item.SD	Item.Tot.Cor	Item.Rem.Alpha	N.Cases
T4	8.0	5.8	2.4	0.38	0.30	137
T5	8.7	7.2	2.7	0.31	0.33	137
T8	7.2	10.9	3.3	0.12	0.45	137
T9	6.3	6.8	2.6	0.11	0.44	137
T11	4.4	3.8	1.9	0.14	0.42	137
T16	5.9	3.4	1.9	0.17	0.41	137
T17	6.3	7.9	2.8	0.19	0.40	137

Table A.9: Evidence post-test: summary statistics (N=137)

TOTAL.STATISTICS	Cronbachs.Alpha	Standardized.Alpha	Total.Mean	Total.Var	Total.SD	N.Items
TOTAL	0.43	0.45	47	73	8.5	7

Other appendices

can be found at the following URLs.

For the DVI version

- The evidence pre-test
- The evidence post-test
- The evidence test coding
- Extract from the Forces teaching package
- Questions from the Forces examination
- The investigation mark scheme
- The issue task

The full hyperlinks

- The evidence pre-test:
<http://www.dur.ac.uk/resources/education/Theevidencepre-test.pdf>
- The evidence post-test:
<http://www.dur.ac.uk/resources/education/Theevidencepost-test.pdf>
- The evidence test coding:
<http://www.dur.ac.uk/resources/education/Evidencetestcoding.pdf>
- Extract from the Forces teaching package:
<http://www.dur.ac.uk/resources/education/ExtractfromSciencetextforces.pdf>
- Questions from the Forces examination:
<http://www.dur.ac.uk/resources/education/Substantiveexam.pdf>
- The investigation mark scheme:
<http://www.dur.ac.uk/resources/education/Investigationmarksheets.pdf>
- The issue task:
<http://www.dur.ac.uk/resources/education/Issuetask.pdf>

Bibliography

- F Abd-El-Khalick, S Boujaoude, R Duschl, N G. Lederman, R Mamlok-Naaman, A Hofstein, M Niaz, D Treagust, and H Tuan. Inquiry in science education: International perspectives. *Science Education*, 88(3):397–419, 2004.
- P. Adey. The case results: implications for science teaching. *International Journal of Science Education*, 14(2):137–46, 1992.
- A.J Baroody, M Cibulskis, M-l Lai, and X Li. Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning*, 6(2):227–260, 2004.
- G.P. Baxter, R.J. Shavelson, S.R. Goldman, and J. Pine. Evaluation of procedure-based scoring for hands-on science assessment. *Journal of Educational Measurement*, 29(1):1–17, 1992.
- T. Bryce, J. McCall, J. MacGregor, I. Robertson, and R. Weston. *TAPS: Techniques for the Assessment of Practical Skills in Foundation Science, Teachers' Guide*. Heinemann, London, 1983. TAPS 1,2,and3 are resource packs that accompany this Teacher Guide.
- A. A. S. Buffler and F Lubben. The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education*, 23(11):1137– 1156, 2001.
- Zhe Chen and David Klahr. All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5):1098–1120, 1999.
- C. A. Chinn and B. A. Malhotra. Childrens' responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, 94:327–43, 2002.
- Barry Cooper. Applying ragin's crisp and fuzzy set qca to large datasets: social class and educational achievement in the national child development study. *Sociology Research On-line*, 10(2), 2005.
- Western Australia Curriculum Council. Science learning area statement. <http://www.curriculum.wa.edu.au/files/pdf/science.pdf>, 1998.
- B. C Davis. *GASP: Graded Assessment in Science Project*. London, Hutchinson, 1989.
- J. F. Donnelly. Secondary science teaching under the national curriculum. *School Science Review*, 81: 27–35, 2000.
- J. F. Donnelly, A. S. Buchan, E. W. Jenkins, and A.G. Welford. *Investigations in science education policy. Sc1 in the National Curriculum for England and Wales*. University of Leeds, Leeds, 1994.
- S. Duggan and R. Gott. Intermediate general national qualifications (gnvq) science: a missed opportunity for a focus on procedural understanding? *Research in Science and Technological Education*, 18(2): 201–214, 2000.
- S. Duggan and R. Gott. What sort of science education do we really need? *International Journal of Science Education*, 24(7):661–679, 2002. Argument based on evidence.
- S. Duggan, K. Foulds, and R. Gott. What makes one investigation more difficult than another? In *British Educational REsearch Association*, Stirling, 1992.

- R.A. Duschl, H.A. Schweingruber, and A.W. (editors) Shouse. Taking science to school: Learning and teaching science in grades k-8. Technical report, Committee on Science Learning, Kindergarten Through Eighth Grade. Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press., 2006.
- Harry B.G. Ganzeboom and Donald J. Treiman. International stratification and mobility file: Conversion tools. Internet, 2005. URL <http://www.fsw.vu.nl/~h.ganzeboom/ismf>.
- P.J Germann and R Aram. Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of Research in Science Teaching*, 33(7): 773–798, 1996.
- P.J Germann, R Aram, and G Burke. Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, 33(1):79–99, 1996.
- J. Glaesser, R. Gott, R. Roberts, and B. Cooper. Success in open-ended investigations in science: using qualitative comparative analysis to identify necessary and sufficient conditions. *Research in Science and Technological Education*, 2009.
- R Gott and S Duggan. Problems with assessment of performance in practical science: which way now? *Cambridge Journal of Education*, 32(2):183–201, 2002. Argument based on evidence.
- R Gott and S Duggan. A framework for practical work in science and scientific literacy through argumentation. *Research in Science and Technological Education*, 25(3):271–291, 2007.
- R Gott and S Duggan. *Understanding and Using Scientific Evidence*. London, Sage, 2003a.
- R Gott and S Duggan. Science investigations in a virtual laboratory: Building success in sc1 science. CD ROM, 2003b.
- R. Gott and P. Johnson. Science in schools: Time to pause for thought? *School Science Review*, 81 (295):21–28, 1999.
- R Gott and P Murphy. *Assessing investigations at ages 13 and 15. Science report for teachers:9*. DES WO, 1987.
- R. Gott and R Roberts. A written test for procedural understanding: a way forward for assessment in the uk science curriculum? *Research in Science and Technological Education*, 22(1):5–21, 2004.
- R Gott, K Foulds, P Johnson, M Jones, and R Roberts. *Science Investigations 1*. London, Collins Educational, 1997.
- R Gott, K Foulds, R Roberts, M Jones, and P. M Johnson. *Science Investigations 2*. London, Collins Educational, 1998.
- R Gott, S Duggan, and P. M Johnson. What do practising and applied scientists do and what are the implications for science education? *Research in Science and Technological Education*, 17(1):97–107, 1999a.
- R Gott, K Foulds, R Roberts, M Jones, and P. M Johnson. *Science Investigations 3*. London, Collins Educational, 1999b.
- A. Gotwals and N. Songer. Measuring students’ scientific content and inquiry reasoning. In S.A. Barab, K.E. Hay, N.B. Songer, and D.T. Hickey, editors, *Making a Difference: The Proceedings of the Seventh International Conference of the Learning Sciences (Icls)*. International Society of the Learning Science, 2006.
- M Haigh. *Investigative Practical Work in Year 12 Biology Programmes*. Unpublished PhD thesis, University of Waikato, 1999.

- C. Hart, P. Mulhall, A. Berry, J. Loughran, and R. Gunstone. What is the purpose of this experiment? or can students learn something from doing experiments? *Journal of Research in Science Teaching*, 37(7):655–675, 2000.
- D. Hodson. Practical work in science: time for a reappraisal. *Studies in Science Education*, 19:175–184, 1991.
- B. Inhelder and J. Piaget. *The growth of logical thinking*. Routledge and Kegan Paul, London, 1958.
- E. W. Jenkins. *From Armstrong to Nuffield: studies in twentieth century science education in England and Wales*. John Murray, London, 1979.
- M. Jones and R. Gott. Cognitive acceleration through science education: alternative perspectives. *International Journal of Science Education*, 20(7):755–768, 1998.
- David Klahr and Milena Nigam. The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10):661–667, 2004.
- D. Kuhn and D. Dean. Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11):886–870, 2005.
- D. Kuhn and E. Phelps. The development of problem-solving strategies. *Advances in Child Development and Behaviour*, 17:1–44, 1982.
- D. Kuhn, E. Amsel, and M. O’Loughlin. *The development of scientific thinking skills*. Academic press, 1988.
- D. Kuhn, L. Schauble, and M. Garcia-Mila. Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9:285–327, 1992.
- Z. Kunda. The case for motivated reasoning. *Psychological Bulletin*, 108:480–498, 1990.
- J.R. Martin. *Writing science: literacy and discursive power*. University of Pittsburgh Press, Pittsburgh, 1993.
- R. Millar, R. Gott, F. Lubben, and S. Duggan. The packs project: full report of research activities and results. Technical report, Universities of York and Durham, 1995.
- J.J. Mintzes, J.H. Wandersee, and J.D. Novak. *Teaching science for understanding; a human constructivist view*. Educational Psychology Series. Academic press, 1997.
- J. Osborne, S. Simon, and S. Collins. Attitudes towards science: a review of the literature and its implications. *Journal of Research in Science Teaching*, 39:1049–1079, 2003.
- Michael Polanyi. *The Tacit Dimension*. 1966.
- QCA. Science: The national curriculum for england? Technical report, Qualifications and Curriculum authority, 2004.
- Qualifications and Curriculum Authority. Gcse criteria for science, undated. Specification for GCSEs from 2006.
- Charles C Ragin. *Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA)*, volume Forthcoming in Benoit RihConfigurational Comparative Analysis, chapter 5. Sage Publications, 2007.
- R Roberts. Using different types of practical within a problem solving model of science. *School Science Review*, 85(312):113–119, 2004.
- R Roberts and R Gott. The role of evidence in the new ks4 national curriculum for england and the aqa specifications. *School Science Review*, 87(321):29–40, 2006a.
- R. Roberts and R. Gott. Procedural understanding in biology: how is it characterised in texts? *School Science Review*, 82(298):83–91, 2000.

- R Roberts and R. Gott. Assessment of biology investigations. *Journal of Biological Education*, 37(3): 114–121, 2003.
- R. Roberts and R. Gott. Procedural understanding: its place in the biology curriculum. *School Science Review*, 81:294–, 1999.
- R Roberts and R Gott. Assessment of performance in practical science and pupil attributes. *Assessment in Education*, 13(1):45–67, 2006b.
- R Roberts and R. Gott. Evidence, investigations scientific literacy: what are the curriculum implications? In *Proceedings of NARST conference, New Orleans*, April 2007.
- J Ryder and J Leach. University science students' experiences of investigative project work and their images of science. *International Journal of Science Education*, 21(9):945–956, 1999.
- Jim Ryder. Identifying science understanding for functional scientific literacy. *Studies in Science Education*, 36:1–44, 2001.
- W.A. Sandoval. Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1):5–51, 2003.
- W.A. Sandoval and B.J. Reiser. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3):345–372, 2004.
- L Schauble. The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1):102–119, 1996.
- P Screen. A case for a process approach: the warwick experience. *Physics Education*, 23:146–49, 1988.
- M. Shayer and P.S. Adey. Accelerating the development of formal thinking in middle and high school students, ii: post project effects on science achievement. *Journal of Research in Science Teaching*, 29(1):81–92, 1992.
- G. Solano-Flores and R.J. Shavelson. Development of performance assessments in science: conceptual, practical and logistical issues. *Educational measurement: issues and practice*, 16(Fall 3):16–25, 1997.
- G. Solano-Flores, J. Jovanovic, R.J. Shavelson, and M. Bachman. On the development and evaluation of a shell for generating science performance assessments. *International Journal of Science Education*, 21(3):293–315, 1999.
- Science The House of Commons and Technology Committee. *Science education from 14 to 19*. Number 1. London, The Stationery Office, third report of session 2001-2 edition, 2002.
- K.A. Toh and B.E. Woolnough. Science process skills: are they generalisable? *Research in Science and Technological Education*, 12(1):31–42, 1994.
- E. Toth, D Klahr, and Z Chen. Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, 18(4): 423–459, 2000.
- R. Tytler, S. Duggan, and R Gott. Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education*, 23(8):815–832, 2001a. Empirical.
- R. Tytler, S. Duggan, and R Gott. Public participation in an environmental dispute: implications for science education. *Public Understanding of Science*, 10(4):343–364, 2001b. Empirical ISI Document Delivery No.: 506FA.
- R. Watson, A. Goldsworthy, and V. Wood-Robinson. What is not fair with investigations? *School Science Review*, 80(292):101–106, 1999.
- A.G. Welford, W. Harlen, and B Schofield. *Practical testing at Ages 11, 13 and 15*. Department of Education and Science, London, 1985.
- B. E. Woolnough. *Practical science*, chapter Setting the scene., pages 16–17. Buckingham, Open University Press, 1991.