Proceedings

Experimental design – can it be taught or learned?

John Garratt and Jane Tomlinson

Department of Chemistry, University of York, Heslington, York, YO10 5DD
E-mail: cjg2@york.ac.uk and jlt7@york.ac.uk

Introduction

Do we teach our chemists the skills they need? This symposium seeks to address this question. We fear that the position has not changed much since one of us concluded that "we should put less emphasis on the teaching of chemistry and more emphasis on learning how to be chemists; because being a chemist involves knowing chemistry, but knowing chemistry does not make you a chemist."[1]
In the context of this symposium we would also add that it really doesn't matter who is asking the question, since learning to be a chemist is one of the best ways of developing the skills needed for almost any role in life.

One of the reasons for this is found in Nyholm's phrase of 'learning **through** chemistry'. A particular benefit of a scientific education is that it provides opportunities to learn to approach problems in a scientific way. What this means is discussed more by philosophers than by scientists. Black, for example, in his book *Critical Thinking*,[2] argued that there **is** something that is properly described as the scientific method, but recognised that it has never been satisfactorily defined. Medawar was one of the few practising scientists who said anything useful about the scientific method. Amongst other perceptive comments, he wrote, "Science, broadly considered, is incomparably the most successful enterprise human beings have ever engaged upon; yet the methodology that has presumably made it so, when propounded by learned laymen, is not attended to by scientists, and when propounded by scientists is a misrepresentation of what they do."[3] In spite of this rather negative comment, he later concluded that "even if it were never possible to formulate **the** scientific method, scientific methodology, as a discipline, would still have a number of distinctive and important functions to perform." We agree with the view that there may be no such thing as **the** scientific method, and accordingly we offer the following definition: "Scientific method consists of an amalgam of generic thinking skills combined and weighted appropriately to reflect the ethos of a particular discipline".[4] This definition indicates our belief that there is no **single** approach to investigations which can be described as **the** scientific method, and that the details of the scientific approach depend on the context. However, there is no doubt that an ability to handle experimental error is an important part of at least some aspects of the scientific approach to investigations. We also propose one universal principle of scientific method; it is that 'Doing an experiment is the last resort of the scientist who has nothing left to think about'. We will try to justify this in posing, as our own version of the title of the symposium, the question 'Do we teach chemists enough about the methodology of science?'

Misconceptions with the language of error

The chemistry Benchmarking Document[5] gives as one of the Practical-Related Skills which chemistry graduates are expected to acquire "the ability to interpret data derived from laboratory observations and measurements in terms of their significance and the theory underlying them".

We have become aware that many first-year chemistry students have misconceptions that would be a severe barrier to the development of these skills.[6] We asked first-year students, as part of their lab report, to
"Write a paragraph summarising the reasons for drawing a straight line through data using an objective rather than a subjective method."
Rather more than half of our 65 respondents gave as a reason for using an objective method (such as least mean squares regression) that it would increase the **accuracy** of their results. With hindsight we can see that this misconception almost certainly arose from the conventional use of the phrase 'line of best fit'. For a student drilled to accept the importance of accuracy, it seemed natural to associate 'best' with 'most accurate'.

In an attempt to investigate the extent of these misconceptions, we asked our first-year students to provide written answers to a set of questions. Our conclusions have recently been published.[7] The questions we asked included the following:

1. An analytical procedure needs to be *precise* and *accurate*. How would you investigate how well a procedure meets these criteria?

2. Under what circumstances would you describe a difference between two values as *significant*?

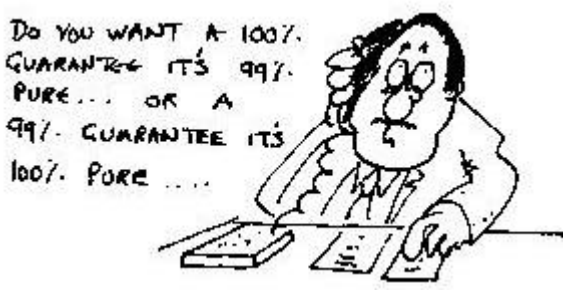
3. Can a *qualitative* procedure prove that a constituent is absent from a substance?

In the written preamble to the questions we made the point that
"The purpose of this exercise is to give you an opportunity to think about and explain or describe how **you** would use, in a scientific context, some words which have both a technical and a general meaning. Remember that the questions are asking what **you** think; they are not asking for the 'correct' answer; (in a sense there is **no** single correct answer, since the meaning may vary with the context)."
This message was reinforced in a short explanatory talk.

We assigned all the responses to one of two categories: those that showed 'some or good understanding', and those that showed 'little or no understanding'. We tried to be generous with our evaluation, and in particular to give credit to responses that showed some understanding, even though they did not meet the requirement of describing how each respondent would **use** the words in question. In spite of our wish to be generous, when we looked at the answers to questions on the investigation of accuracy and precision, we were only able to assign 'some or good understanding' to well under half the responses.

Writing about significant differences, most respondents mentioned the size of the difference as being important, but none gave any indication that they understood that high levels of variation between replicate values (low precision) makes it hard to detect differences in mean values.

Figure 1



When it came to qualitative procedures, only 18% recognised the limitation that you cannot prove that something is **absent**, but only that it is below the level of detection. Lawrence's cartoon (**Figure 1**) taken from *A Question of Chemistry*[8] makes the point succinctly and memorably.

As we have described,[7] the student responses to these and other questions confirmed our view that first-year chemistry students would benefit from a considerably better understanding of the language used to deal with error and uncertainty in quantitative measurement. It may be that by the time the students graduate they will have picked up a good understanding of the language and the procedures, but we fear that they do not have much opportunity to do so. We have little confidence that general textbooks covering this topic do so in a way which deals with the problems faced by students trying to understand how to treat error and uncertainty. Take the word 'accuracy', for example; most books **define** it as something like 'closeness to the true value'. Of course that is what it **means**, but as an explanation it comes close to what Coldstream has called "colluding in a spoon-feeding process".[9] As a definition it is perfectly acceptable for all those students who are still living in Stage 1 of Perry's stages of intellectual development,[10] which has been paraphrased as 'Right answers to everything exist, and these are known to authority whose role it is to teach them'. The definition completely misses the point that, if you know what the true value is, you do not need to measure it. Given that you measure something in order to establish what the answer is, the important question is 'How can you know whether your result is accurate?' So, knowing what accuracy **means** is only the first (and tiny) step in being able to use the word effectively. Similar criticisms apply to other textbook definitions; they are not helpful in an operational world. Furthermore, anecdotal evidence suggests that there is no consensus amongst academics either about the correct usage of words and concepts to describe uncertainty in data, or in the best procedures available for interpreting experimental data. We are thus led to the conclusion that there is a need for much careful thought about the best ways to meet the Benchmark objective relating to data interpretation.

Scientific method in the design of investigations

The Benchmarking document also includes as one of the skills needed by graduate chemists "competence in planning, design, and execution of practical investigations". Of course an understanding of error is a key part of this – at least insofar as we are talking about quantitative chemistry, since the planning process involves thinking about the way the data will be processed.

Planning an experimental design that has the best chance of illuminating the research topic is one of the many things one has to do before the last resort of doing an experiment. But most students only think about errors when they come to write their lab reports **after** they have collected their data. We now report some previously unpublished results that illustrate some of the benefits of not doing an experiment before thinking carefully about the question being investigated and about the best way to collect data that is most likely to provide a definitive answer.

Working in conjunction with Millar,[11] we used our computer simulation pendulumLAB; this allows users to investigate the effect on the dependent variable 'time of a pendulum swing' of the independent variables 'length of the string', 'mass of bob', and 'angle to which the bob is raised'. As part of a larger study we invited experienced academics to carry out a simulated investigation using pendulumLAB. The complete study involved school pupils aged about 14 and first-year university students who used pendulumLAB and several other simulations. Here we report the results obtained by the volunteer academic scientists.

Before starting the exercise, all 15 volunteers were asked to predict the effects of the variables. We regard this as good practice even though, in some investigations, there may be too many possible outcomes for any prediction to be useful. We do not accept that it is 'unscientific' to try to predict the likely outcome of an experiment, because prediction is a useful way to focus on possible outcomes and so to plan a strategy that is likely to distinguish between them. Of course, any such prediction must be followed up by observation; otherwise one ends up like Aristotle, whose reliance on theory led him to assert, amongst other silly things, that the semen of youths between puberty and the age of twenty-one is "devoid of fecundity".[12] We suggest that thinking about likely or possible outcomes of experiments facilitates the rigorous testing of those predictions, that this rigorous testing is the true mark of the scientist, and that the need to do this thinking is another reason why doing an experiment is the last resort of the scientist who has nothing left to think about.

Table 1 Predicted effects of the variables on the time of the pendulum swing as made by 15 subjects

Predicted effect of increase in	Length	Mass of bob	Angle
Increase	13	7	4
No effect	1	6	10
No prediction	1	2	1

Table 1 summarises the predictions made by our 15 volunteer subjects. Before presenting the data these subjects collected, we will consider how these predictions might be tested rigorously and efficiently. The first rational step in investigating the nature of any effect would be to test whether or not an effect is observable; there is no point in trying to establish an exact relationship if no effect can be demonstrated. In establishing whether or not an effect can be measured, it is worth remembering the principle of falsification as propounded by Popper (see, for example, ref. 3). According to this principle, an hypothesis is useful when it is framed in such a way that it can be **disproved**. It follows that the **prediction** of a positive effect (such as 'the mass of the bob **does** have an effect on the time of the swing'*)* does not translate directly into a useful **hypothesis** because it cannot be disproved; it is possible to show that any effect is too small to be measured using the available procedure, but it is not possible to prove that there is **no** effect. It is relevant to recall that the philosophical impossibility of proving the absence of a substance (or an effect) was not appreciated by most of our first-year students.

In contrast to the impossibility of disproving a prediction of a positive effect, any hypothesis that there is no effect is disproved if an effect is actually observed. Thus the prediction that angle or mass has **no** effect is an hypothesis in Popper's sense. An efficient way to test either hypothesis is to hold two variables constant, pick two values of the third which are as far apart as is reasonable, and make enough measurements at each of these values to be able to carry out a valid statistical test of the difference between the mean values. This involves an underlying assumption that any effect is always in the same direction, but it is nevertheless a useful starting point. It is also relevant to recall that the problem of detecting a significant difference between two variables is another of the concepts with which our first-year chemists seemed to be unfamiliar.

Two predictions are of special interest to the analysis of the strategy used by our volunteers. These are the prediction that angle has no effect on the time of swing (predicted by ten subjects) and that the mass of the bob does have an effect (predicted by seven subjects). Thirteen of our fifteen subjects came into one or both of these categories. Both these predictions are wrong and are interesting for different reasons. Angle actually does have an effect (though it is very small at low angles). Thus this prediction can rather easily be proved wrong, and so those who made it might be expected to change their minds as a result of doing

the experiment. In contrast the mass of the bob has no effect (at least not at a level which has ever been detectable with the most sophisticated equipment). Thus this prediction cannot be falsified. Failure to observe an effect need not lead to the conclusion that the prediction is wrong, since it would be legitimate to conclude that the predicted effect was too small to be detected. In practice, it would be hard for a rational scientist to persist with a theory in the absence of any positive evidence on the grounds that a predicted effect was too small to be detected. However, one would hope and expect that they would only change their minds after a thorough investigation.

Table 2 Conclusions on two selected predictions after carrying out the investigation

	Confirmed by investigation	Left uncertain by investigation	Changed mind after investigation
Predicted no effect of angle	7	2	1
Predicted effect of mass	1		6

Examples of simulated investigations

Table 2 shows that our subjects did not respond as described above, and that only one subject (out of ten) was convinced of the correct conclusion that angle has an effect, whereas six (out of seven) rejected their original prediction by concluding that the mass of the bob has **no** effect. Inspection of the data collected by these subjects shows that their investigations were not carried out according to the principles outlined above, and that this may explain the somewhat paradoxical conclusions they drew.

Considering first the effect of angle, we found that seven of the ten used a strategy that made it difficult to refute the hypothesis. Six of them took either no replicate readings, or made only one duplicate or triplicate measurement. Three of these six took five or fewer measurements. Four of the seven (one of whom did take replicate readings) used a range limited to 30 degrees or less. Although it was one of this group whose opinion changed as a result of the investigation, it is plausible that most of them viewed the investigation as an opportunity to **confirm** their prediction, rather than to **disprove** an hypothesis.

The remaining three subjects in this group of ten made between twenty-two and seventy-two relevant measurements and, importantly, made four to six replicate measurements at more than one angle. A small selection of the data they collected is shown in Table 3. Subject 7948 concluded that "time is independent of mass and angle". The other two concluded that there might be an effect. Subject 170 wrote: "The angle of swing has no (or very little) effect on the time for 10 swings. But there appears to be a **slight** decrease in time for swings with decreasing angle, which does not seem entirely within experimental error", and 7584 wrote "Time increases a little bit with the angle, but this may very well be due to experimental error." The data in Table 3 have been deliberately selected from each subject's total set to illustrate how easy it is to demonstrate a positive effect. The data from subject 7584 show that the effect is harder to see when the length of the string makes for a short time of swing. But even the results selected from this subject provide convincing evidence of a significant difference between the two chosen angles. Two reasons can be suggested for these subjects not recognising the effect. One is that they were so committed to their original prediction that they did not look critically at their evidence (hardly the mark of an objective scientist). The other is that the evidence was obscured by the way it was presented by the computer. The software allows them to view all the data collected, but it lists it in the order in which it is collected. The data shown in Table 3 were not collected in two sequential blocks as displayed in the table, and so the process of abstracting the data from the complete set makes the effect easier to notice. An alternative to abstracting the data is to plot a graph, and the software allows this. However, where an effect is small (as it is in the case of angle) it is much easier to see it when plotted on paper than when displayed on a computer screen. Both these disadvantages of data presentation were almost certainly factors in obscuring the significance of the results.

Whatever the real reason why these subjects did not change their minds as a result of carrying out their investigation, we suggest that, even though they took replicate measurements, they were guilty of doing experiments while they still had things to think about.

Turning to the seven subjects who predicted that mass of the bob would have an effect we see that six of them actually changed their minds, by preferring the conclusion that there is no effect to the conclusion that the effect is too small to be measured. This is surprising, given that the **absence** of an effect is virtually impossible to prove, and our

Table 3 Selected data from three subjects who did not identify a definite effect of angle after carrying out the investigation

Subject number	7948		170		7584	
Fixed Variables	L	M	L	M	L	M
	100 cm	100 g	100 cm	50 g	5 cm	100 g
Angle	20^o	80^o	1^o	90^o	10^o	80^o
Conclusion	**No effect**		**Possible effect**		**Possible effect**	
Readings	20.1	22.4	19.9	23.8	4.5	5.4
	20.0	22.7	20.0	23.5	4.3	4.7
	20.3	22.5	20.1	23.2	5.0	5.0
	20.3	22.6	20.5		4.5	4.9
		22.6	20.4		4.3	5.1
			20.1		4.3	
Mean	20.17	22.56	20.33	23.50	4.48	5.02
S.E.M.	0.08	0.05	0.12	0.17	0.11	0.12
Relevant measurements	22		33		72	

subjects' investigations of angle suggest that they are remarkably resistant to changing their minds. Furthermore, these subjects can hardly claim that their conclusion was based on an exhaustive study. The six who changed their minds made between five and twenty-two relevant measurements, and only one of these took more than one replicate measurement. This latter subject first took single measurements at nine masses from 10 to 90 g, and then twelve replicates at a mass of 100 g. From the point of view of an effective strategy, we point out that it is actually harder to make a statistical comparison of several single values with one mean than it is to compare two mean values based on (equal numbers of) replicates. So it seems that these subjects were persuaded to change their minds on the basis of a less than rigorous investigation. The one subject who confirmed the initial (incorrect) prediction that there is an effect of mass based this conclusion on only five measurements, again suggesting a tendency to look for confirmation of a prediction rather than trying to disprove an hypothesis.

Conclusions

What conclusions can we draw from the evidence that these well-qualified and experienced scientists used strategies that might be described as naïve when judged against basic criteria associated with a scientific approach? We emphatically do **not** suggest that they do not know how to conduct investigations. It is important to take into account the artificiality of their situation. They were given a short introduction to the project, and then asked to carry out their investigation. They are all busy people and unlikely to give as much considered thought to the problem we set as they would to an investigation of genuine interest to them. It therefore seems reasonable to suggest that they reverted to an intuitive strategy. For almost all of them this involved holding two variables constant, and systematically varying the third. This is a necessary strategy for investigating the nature of a relationship between two variables, but it is not an efficient strategy for establishing whether an effect can be detected, and (as argued above) it seems rational to establish this before spending time and effort in investigating the nature of any relationship. We conclude that the Popperian principle of formulating hypotheses with a view to disproving them is not intuitive and has not been embedded in the subconscious of these subjects. The fact that only six of the fifteen systematically made replicate measurements suggests that this principle is less automatic than one might expect given the emphasis placed on it in most laboratory courses. It is a humbling thought that our complaints about the deficiencies of students are reflected in our own performance when we are placed in an unfamiliar situation. If the scientific method is assumed to be understood intuitively by scientists,[3] then this evidence suggests that intuition might be improved by some formal instruction, and that we should take this into account when we address the question "Do we teach our chemists the skills they need?"

We are convinced that one of the main benefits of a scientific education ought to be that it leads to the development of a deeply ingrained appreciation of some principles of scientific method. We do not believe that these are learned by osmosis from the kind of laboratory course which most of us run.

This means that we need to rethink our laboratory courses with the specific objective of helping students to develop an appreciation of the principles to use in planning investigations, and that this involves including explicit advice on the determination and quantification of errors and on the appropriate ways of planning to take account of uncertainty. Of course most of us investigate much more complex systems than a pendulum. Each system will yield to a different combination of thinking and experimenting. Sometimes it is efficient to do a quick experiment and then do a lot of thinking. Sometimes it is much better to spend a lot of time thinking before embarking on the last resort of an experiment. What determines the optimum strategy? Is it possible to draw up a set of guidelines that will lead one to an optimum strategy? If so, is it possible to devise learning opportunities of direct relevance to chemistry through which these can be learned? As yet we have no clear answers to these questions. However, we believe that the answer is **'yes'**, in spite of Medawar's comment that "…those who have been instructed [in scientific method] perform no better as scientists than those who have not". We therefore suggest that it would be worthwhile for a group of interested individuals to consider both what principles of scientific method should be explicitly taught and what methods of teaching and learning are most likely to be effective. On the basis of such a set of guidelines it should be possible to develop a valuable new range of teaching resources.

We do, however, end with a cautionary note. In the end, in teaching our chemists the skills they need, all we can really do is to stimulate and enthuse them, and point them in the right direction.

Acknowledgements
Different aspects of the work described here were carried out, as indicated, in collaboration with Andrew Horn, Paul Dyson, and Robin Millar. We are grateful to Glaxo Wellcome plc and to the ESRC (grant number RO22 25 0121) for financial support.

References.

1. J. Garratt, *U. Chem. Ed.*, 1997, **1**, 19.
2. M. Black, *Critical Thinking, an Introduction to Logic and the Scientific Method* (2nd edn), Prentice-Hall, Englewood Cliffs, NJ, 1962.
3. P.B. Medawar, "Induction and Intuition in Scientific Thought" in *Pluto's Republic*, Oxford University Press, Oxford, paperback edition 1984.
4. J. Garratt, T. Overton, J. Tomlinson and D. Clow, *Active Learning in HE*, 2000, **1**, 152.
5 *General guidelines for the academic review of Bachelors Honours Degree Courses in Chemistry* 1998, Quality Assurance Agency , Gloucester.
6. J. Garratt, A. Horn and J. Tomlinson, *U. Chem. Ed.*, 2000, **4**, 54.
7 J. Tomlinson, P. J. Dyson and J. Garratt, *U. Chem. Ed.*, 2001, **5**, 16.
8. J. Garratt, T. Overton and T. Threlfall, *A Question of Chemistry*, Longman, London, 1999.
9. P. Coldstream, *U. Chem. Ed.*, 1997, **1**, 15.
10 D. C. Finster, *J. Chem. Ed.*, 1991, **68**, 752.
11. R. Millar, *Understanding how to deal with experimental uncertainty: a 'missing link' in our model of scientific reasoning*?, in M. Komorek et al. (eds.), *Research in Science Education. Past, Present and Future* (pp. 276-278). Proceedings of the Second International Conference of the European Science Education Research Association (ESERA), 31 Aug. - 4 Sept. 1998, Kiel, Germany, 1998.
12. P. and J. Medawar, Aristotle to Zoos, Oxford University Press, 1985.