Quality Analysis of Data in Student Experiments

Students' evaluation of experimental evidence within Key Stage 4 investigations

Simon Elliott, 1st February 1997

In a metaphysical sense, the study of science has Dynamic Quality, constantly changing and intellectually "cutting-edge" however, to be able to meet the challenge of true scientific creativity, skills must first be acquired. These are the basic scientific truths and processes that make up the Static Quality and are required in order to fully understand how to investigate the curiosity in the world around us that is the subject of science. These skills need to be developed through educational processes but not in such a way as to give the impression that real scientific investigation happens in this static way. Unfortunately the way that Sc1 investigations are usually approached as linear processes could give rise to an artificial view of scientific research as it gives the impression that investigating ends with the data obtained and the conclusions drawn from it without any consideration for using that data to analyse the original hypothesis and amend and re-test it as a 'real' scientist would.

One of the impacts of the changes made to the National Curriculum within science after the Dearing report was the emphasis placed on the evaluation of data produced during scientific investigations and this does go some way to producing a dynamic, feedback based approach to investigative science

Of course, this was actually nothing new, students had been ensuring that they were performing a fair test, were producing a range of data and collecting valid and reliable data under the previous version of the curriculum but this was the first time that this had to be evaluated as a discreet section.

For example, looking at a sample Sc1 investigation criteria sheet created for the well-used experiment of the reaction between hydrochloric acid and calcium carbonate in the form of marble chips within the Nuffield syllabus, double award co-ordinated science, it is possible to outline evaluationary statements within the assessment criteria;

Students were asked: "What factors affect the rate of a chemical reaction?"

The following criteria were used to assess the students' work;

Level 7

Strand (i) Suggest, with reasons, the factors affecting the rate of reaction

Strand (ii) Obtain results for different sizes of marble chips reacting at varying concentrations and/or temperatures

Strand (iii) Draw conclusions from data obtained as to the effect of factors on the rate of reaction and their relative importance

Level 8

Strand (i) Give a quantitative prediction linking two of the factors mentioned above to the rate of reaction and comment on their relative effects

Strand (ii) As above, using appropriate measuring devises and units (i.e., time to 1 sec, etc.)

Strand (iii) ***Explain why only one factor was varied in each series of experiments whilst all other factors were kept constant - the fair test***

Level 9

Strand (i) ***Suggest repeating the experiment to improve accuracy of results to ensure valid and reliable data.***

Strand (ii) Carry out the above procedures in a systematic manner using two or more techniques to collect data (including secondary sources.).

Strand (iii) Use data obtained to draw series of graphs to show how factors affect the rate of reaction. ***Comment upon the accuracy of the data obtained***

However, it should be noted that this analysis of the quality of data was regarded as suited to investigations at levels 8 or above. Below this, students merely had to demonstrate that the experiment was conducted in a fair way and that they had collected a good range of results, both skills that can be easily taught to most students.

When we look at the new marking scheme taken from the revised Nuffield syllabus, again for double award co-ordinated science, we see a new marking section, the "strand (iv)" where students must comment on the quality of data at all but the simplest level.

Skill Area E: Evaluating Evidence

Marks 2
POAE E.2a
Criteria for awarding marks Make relevant comment about the procedure used or the evidence obtained

Marks 4
POAE E.4a
Criteria for awarding marks Comment on the *accuracy* of the observations or measurements, recognising any anomalous results

POAE E.4b
Criteria for awarding marks Comment on the suitability of the procedure and, where appropriate, suggest changes to improve the *reliability of the evidence*
Mark 6

POAE E.6a
Criteria for awarding marks Comment on the *reliability of the evidence*, accounting for any *anomalous results*, or explain whether the evidence is sufficient to support a firm conclusion

POAE E.6b
Criteria for awarding marks Propose improvements, or further work, to provide additional evidence for the conclusion, or to extend the enquiry

Overall level achieved in skill area E

We must therefore, as teachers, be prepared to assist all of our students in acquiring the skill of evaluation of the quality of data and not just those working at level 8 or above under the former scheme.

This is in line with the National Curriculum where, by level 3, students are expected to understand the "fairness" of a task or investigation and by level 5 they should be able to evaluate the data gathered in a simple way.

Before we can do this, we must understand what is meant by scientific evaluation and find out where our students are coming from. This is especially urgent for those currently in year 10 who have had many years of investigating under the former three- strand model but who are studying for a GCSE under the new four-strand one.

Christofi (1988) considers that *evaluation* is the highest order of six skill areas within the cognitive objectives of any science course. He goes on to say that evaluation involves making value judgements about materials and ideas for given purposes and requires the skills of knowledge, comprehension, application, analysis and synthesis to be first mastered. Whilst this would seem to over-complicate data-evaluation, it certainly suggests that students need to be taught the process systematically. As these concepts involve the manipulation of ideas in the mind of the student, *evaluation of data* would be firmly placed within Piaget's Formal-Operational Period' of cognitive development which, according to testing of the students within the school being used for this work using the Science Reasoning Tasks, many students have not even come close to entering at the age of twelve. *Are we expecting too much in year 10?*

Although not now highly regarded as a method for science education because it separated science into a number of discreet skills and argued that the content was unimportant, Process Science recognised that certain areas of science were skills that could be learnt regardless of, and independent of, the content being covered at the time, although this seemed to translate into a course practically devoid of content. The publication *Science in Process* saw these skills as including "*applying, interpreting, classifying, investigating, evaluating, observing, experimenting, predicting, hypothesising, questioning and inferring"* as did many other approaches including Warwick Process Science and the Nuffield approach itself. Unfortunately, the overall approach the "Process Science" courses brought to science education was "criticised as producing an unhelpful and misleading image of scientific enquiry" by many.

Whilst our approach has changed, there can be no doubt that the process of data evaluation is a skill rather than a "fact" and so is most easily learnt through "doing" and most easily assessed in the same way.

In the assessment system at GCSE level before the National Curriculum came into force and introduced Sc1 investigations, Nuffield Co-ordinated Science assessed practical investigations using discreet 'skill areas'. Skill C, the handling of experimental observations and data, expected students to be able to recognise and comment upon anomalous results, deduce patterns in the data and comment upon possible sources of error. This focus on the quality of the data {Quality of Data} rather than the process of producing the data itself {Quality of Method} seems close to 'real' investigative science and perhaps the skills-led focus did make expectations of students clearer.

These acquired skills are important for everyday life but also for those students who go on to study sciences at a higher level. The ability to recognise errors in data is an important skill in ‘A’ level science as “whenever we make measurements using instruments, there is some error or uncertainty in the result, sometimes due to the experimenter and sometimes the equipment”. This is certainly not a skill that can be taught without a good understanding of what Quality of Data means, built up over a number of years.

I was unable to find any references to research into students’ understanding of data evaluation and so, in order to understand where students are coming from, I decided to use a sample investigation, missing only the conclusion and evaluation, with a group of thirty year 10 students of average ability and ask them to complete the write-up as best they can, together with a number of questions probing their thoughts on what “evaluation” means. The replies to the investigation and questions were then gathered together and collated into groups of similar answers.

First I will look at the replies given for the investigation (a summary of the write-up given to the students is shown below).

A report on an investigation into the motion of pendulums.

Aim:

To discover what factors affect the speed at which a pendulum swings.

Prediction:

I think that a pendulum will swing faster the shorter it is because the bob has to travel a shorter distance to complete one swing because the time taken for one swing will be proportional to the length of the path described by the bob which can be calculated by;

Path length, $P = (\eta / 360) \times (2 \times \pi \times L)$ where $\pi = 3.14159265359$

If this model is correct then I predict that a pendulum with twice the path length of another will take twice as long to perform one swing provided that the angle described by the swing is kept constant.

Method:

I will need to choose pendulum lengths that double each time to make testing the model easier and I will need to make sure that I pull the pendulum back by the correct angle each time. To do this I will make a cardboard template of a 60^{0} segment and attach it below the fixing point for the pendulum. To time one swing accurately would be difficult so I will time twenty and divide that time by twenty to get the time for one swing.

Results:

Here are the results that I gathered from my experiment;

Length, L (cm)	10	20	40	80	160
Path, P (cm)	10.472	20.944	41.888	83.776	167.552
Time for 20 (s)	17.420	22.080	29.260	37.900	46.940
Time for 1 (s)	0.871	1.104	1.463	1.895	2.347
Model time (s)	0.871	1.742	3.484	6.968	13.936

(The question sheet also included a graph of the model results and the real results.)

Conclusion:

The students had around half a side of A4 paper to fill in their answer.

Evaluation:

The students had around half a side of A4 paper to fill in their answer.

The document given to the sample of thirty students had a Flesch reading ease score of 78.3 giving it an approximate reading age of 12 years of age. This means that, although conceptually quite difficult to complete (as is any investigative report), the language used should not have significantly affected the results. However, with a small sample such as the one used, a wider sample should be taken and re-assessed if the conclusions drawn later are to be substantiated.

The students sampled would (or should!) be familiar with the concept of a theoretical 'model' being the result of a prediction from previous investigative work done over their preceding years at the school.

Another possible factor involved is the quality and method of the teaching of scientific investigation at the subject school and further research over a number of different schools would also be needed to be more confident in the results.

<u>Summary of student responses:</u>

General theme of reply with the percentage of replies (n=30)

Predicted model was wrong but the experiment was enjoyable : 3.2%
e.g.; 'I found that the real results were very different from the model so my model was wrong, but it was a good experiment to do because it was quick'

Predicted model appeared to be wrong but maybe it was the data that was wrong : 12.9%
e.g.; 'the prediction was wrong......I think that some of the results could be wrong'

Everything was fine about the experiment, evaluation agreed : 25.8%
e.g.; 'I think that my prediction was more or less right...my results were accurate... fair test''

Everything was fine about the experiment, no evaluation at all : 3.2%
e.g.; 'I can see from the results that the predicted model was correct'

Predicted model was close to reality, repetition of data would help proof {Quality of Data} : 29%
e.g.; 'I was right to say that the pendulum would swing faster but I need to make my results more accurate'

Predicted model was close to reality, changing the method would help {Quality of Method} : 25.8%
e.g.; 'From the graph I can see that my prediction was almost right ...... but I think the experiment could be improved by using a greater range of lengths'

As you can see from the above summary of responses, only around 42% of replies mentioned "data" explicitly within their reply and 83% reported the predicted results {model} to be close to those seen {reality} even though they were vastly different and shown to be as such by the graph included. This appears to be a common feature amongst students, even at KS4, where they seem to find it difficult to abandon their model even when their results clearly disprove it. This was backed up by the work of Kuhn *et al* who, in their investigations into how subjects co-ordinated theory and evidence, noted "instances of adjusting theories to fit the evidence."

This difficulty in challenging their own scientific preconceptions could arise from the very linear definition of investigative science promoted by the curriculum. Within the structure that exists, students hypothesise, test and analyse the data produced. "Real" science is a far more complex process and one that is normally cyclical;

A cyclical view of Sc1: If you approach investigations from the linear model then you arrive at a negative result most of the time. If the sum total of a students work is "this investigation showed that my predictions were wrong" then the student is left with a feeling of failure. When a cyclical approach is used, the student can then say "at first I found that my results did not agree with my prediction and so I looked at where my theory could be improved to come closer to what I observed". This is a far more positive approach and one that more closely matches how scientists operate in the "real" world. By giving students a piece of work where they are likely to find that they are wrong and cannot do anything about it is likely to lead to the results seen in the survey above where they cope with this by not admitting the difference between their hypothesis and reality. This is all too often repeated in their own complete investigations when they seem unwilling to accept the errors present and instead adjust their conclusions, and accepted scientific theories, to fit their findings.

In the second part of the survey, students were asked to respond to three statements and I will look at their answers to each question and the reasoning behind it in turn.

<u>Summary of responses to question one</u>

"**What do you think it means to evaluate an experiment"**

With this question I was looking to see whether the students would focus on the data produced in an investigation or experiment or on the method of the investigation or experiment itself. In analysing the responses, Quality of Data was taken to be indicated by comments such as "justify any irregularities in the results" and Quality of Method indicated by comments such as "say what you did and how you would do it better". The question itself was deliberately vague so as not to prompt them down one route or the other. This may have had the effect of confusing the students but further investigation would have to be done to confirm or deny this.

General theme of responses with the percentage of replies (n=30)

How the experiment went and what you could do to improve it {Quality of Method} : 33%
e.g.; 'it means to say what you did and why you did it'

Overall quality of the data produced {Non-analytical Quality of Data} : 6%
e.g.; 'to say how good the results were'

Fairness of the experiment {The Fair Test} : 20%
e.g.; 'whether the experiment was a fair test'

How much the experiment has taught us {Quality of Learning Outcomes} : 10%
e.g.; 'to say what you have found out from the experiment'

Overall quality of the data produced {Analytical Quality of Data} : 26%
e.g.; 'to justify any irregularities in the results'

Do not know : 5%

The picture from the replies to this question showed that one-third of the students think that to evaluate a scientific investigation is the same as would be expected in the evaluation of a project within design technology, that is a judgement on the quality of the way that the experiment was conducted - the quality of the method itself - and not of the data produced. There is obviously a need to explain the difference to the students. Only 26% of the students' replies considered that evaluation was about making judgements on the reliability and accuracy of the data produced. This is clearly worrying as this would limit the students to 2 marks in the evaluative strand in a KS4 investigation to the other 74% if they used the same model as their above reply when writing up their report.

The reason for the observed evidence became clear when looking at the document 'Technology in the National Curriculum' in which students have been expected to evaluate their work *as a separately assessed attainment target* for far longer than science. For example, at level one, students are expected to *describe to others what they have done and how well they have done it* and at level three, *comment on ....how the task was tackled*. These two examples and many of the others, focus on the Quality of Method completely. It is not therefore surprising that our student in science produce statements of this sort when asked to evaluate an investigation. Perhaps we need to re-appraise our approach to strand iv of Sc1 completely if we are to remove the students' linking of evaluation to their experience in the technology curriculum area when they are working within science. Certainly further work needs to be done in this area.

There are two main reasons for worrying if students have failed to grasp the 'true' skill of scientific evaluation. Firstly the simplistic reason of "it's in the National Curriculum and assessed at GCSE by all syllabuses including Nuffield Co-ordinated Science (Double Award) so we had better get it right" and secondly the more important reason that we, as science educators, are trying to instil skills in our students that will serve them in their lives whether they pursue a career in science or not. The ability to not take data on face value is one of the most widely used skills that fall into this category.

The responses to this question would indicate that most students do not focus on the data but on the method of producing it. Whilst this may be a subtle difference, I would say that the perspective that a person approaches a problem with fundamentally affects the way that they deal with the problem and as a result their effectiveness at achieving a successful solution.

<u>Summary of responses to question two</u>

"**If you see something happen in a scientific experiment then it must be right because scientific experiments are always right. Discuss."**

After looking at the previous two questions and contemplating the answers that the students may produce, I wanted to investigate whether they attached some mystical significance to scientific data because of the way it is portrayed by the world around them.

This view of scientific experimentation as producing definite and irrefutable observations that always back up the theory under investigation is all to often reinforced by teachers themselves every day when they perform demonstrations in front of groups to the line of "and what this *proves* is......." instead of "one *possible* explanation for what we have just seen is....". This is commented upon in great length in Driver *et al*(1996) and presented as one of the main reasons that students have problems when their observations do not fit the theory.

General theme of responses with the percentage of replies (n=30)

a. positive responses

If done carefully then there is reasonably certainty : 3.3%
*e.g.; 'scientific experiments **should** be right if they are done carefully*

If a scientist has done the work then you can be certain : 3.3%
e.g.; 'an experiment in a book is always right', 'if a scientist did it, you can be pretty sure'

b. negative responses

Results are different every time so you can only get *average* answers. Likely but not certain : 33%
e.g.; 'things may be out of control of the experimenter so we have to average the results'

Results are so affected by circumstances that they cannot be certain : 52.8%
e.g.; 'nobody can ever do an experiment completely right'

An experiment just produces data, it is the experimenter that makes the value-judgement : 6.6%
e.g.; 'it is not the results that are wrong, its the person doing the test that says if they're OK.'

The replies to this question again fell more into commenting on the uncertainty of the *method* producing the data itself rather than the uncertainty of the data itself. 53% of replies felt that data can *never* be totally correct as mistakes are *always* made when conducting experiments however 33% of replies did comment that the more times the experiment was repeated, the more certain you could be in the results so some appreciation of the quality of data was evident.

What the responses did show was that the groups under investigation do, on the whole, have a healthy distrust of experiments however their focus was the method. They failed to see that, even when the experiment was conducted correctly there could still be errors (Fullick, 1994) that may not necessarily disprove the theory or that the apparently erroneous data could point to a need to *revise* the theory. This could be an indicator of them understanding of science as having Static Quality rather than the Dynamic Quality which I would suggest that it has.

<u>Summary of responses to question three</u>

"**Why do you think that it is necessary for scientists to evaluate the data that they have gathered in an experiment?"**

Here again I was looking to whether the students would approach the question from a Quality of Method perspective rather than a Quality of Data one. This time I changed the focus of the question from one where the nature of the experimenter was ambiguous to one where the label "scientist" was used. All of the responses featured "they" rather than the "we" that had appeared in previous responses and there was a far narrower range of answer types than in previous questions. The stress of the question was also made more specific, using the word *data* itself.

General theme of responses with the percentage of replies (n=30)

To say what the data meant {confusion with conclusion/analysis} : 16%
e.g.; 'because they have to summarise all of the results and what they thought of them.'

To analyse where the *method* went wrong {Quality of Method} : 43%
e.g.; 'they then know what has been done wrong and how they can improve the method'

To judge the quality of the data {Quality of Data} : 40%
e.g.; 'it would help other people judge the reliability of the results'

Even though the stress was placed upon data and yet still only 40% made any comment about judging the quality of the data itself. Stephen Knutton describes students as having the need to "interpret results, evaluate their validity and relate their findings to an appropriate level of scientific understanding." in order to succeed in scientific investigation. The answers to question three show that many will not therefore succeed in the present model of investigations as they do not seem to understand what is fundamentally required. Does the true understanding of evaluation only come when students become scientists and they are investigating to answer a curiosity rather than as a piece of required assessment or do the results say something more fundamental about the way that secondary school students, at the subject school at least, are taught science. Without the curiosity that drives science, the subject soon falls into a pattern of static facts to be learnt. Tyler and Swatton attacked the very concept of the Sc1 investigation as "Sc1 gives us an instrumental view of what it means to be scientific, as if the essence of investigation lies in its component methodological processes. However the reality is altogether richer and more varied. Imposing a set of methodological rules on a process which is above all *creative* can only serve to diminish its potential as a way of learning in science."

Conclusion and Evaluation

The small sample of data used in this study could not be expected to produce conclusive results however I feel that it is clear from the picture in each of the four sections that students no not, for the majority, have the same concept of evaluation that we are trying to develop within them. For whatever reason, they are looking to the method itself rather than the results and this means that, whether for the purposes of examination assessment or life-skills, they are failing to understand the true nature of scientific investigation. As the work of Driver and others would suggest, this is indicative of a closed, static view of science from students lacking that curious urge to really understand the "why" of the world around them. Does the ability to understand the tentative nature of data only come with a deeper understanding of what it means to be a scientist?

References

[i] Pirsig R.M. (1991) Lila, An enquiry into morals, Black Swan

[ii] Christofi, C. (1988) 'Assessment and Profiling in science', Cassell

[iii] Donaldson, M (1978) 'Children's Minds' (appendix), Fontana

[iv] Mussen, P.H. (Ed), Piaget, J (1970) 'Manual of child psychology, Piaget's Theory of Stages', Wiley and sons

[v] Inner London Education Authority {ILEA} (1987) Science in Process, Heinemann

[vi] Screen, P. (1986) An introduction to Warwick Process Science, Ashford Press

[vii] Driver, R. *et al* (1996) Young people's images of science, Open University Press

[viii] Ingram, Neil R. (1990) Assessing Nuffield Co-ordinated Sciences, Longman

[ix] Fullick, A.&P. (1994) Chemistry (Heinemann Advanced Science), Heinemann

[x] Kuhn,D. *et al* (1988) The development of scientific thinking skills, Academic Press

[xi] Wellington, J (1994) Secondary Science, Contemporary Science and Practical Approaches, Routledge

[xii] DFEE (1990) Technology in the National Curriculum, HMSO

[xiii] Knutton, S. (1994) Assessing practical work in science, Routledge

[xiv] Tyler, R. and Swatton, P. 'A critique of AT1 based on case studies of student investigations', School Science Review 74(266):21-35