Modelling evaluation methods in Key Stage 3 scientific enquiries

Lindsay Davison, Bohunt School, Hampshire

July 2015
url:

Keywords: *Science, Improving, Adapting, Evaluating, Teacher modelling, Scientific enquiry, Key Stage 3, Year 7*

Context

Bohunt School is highly successful secondary comprehensive school for students aged 11 to 16 located in the rural setting of Liphook in Hampshire, with more than 1575 students on roll. My role within the school is Head of STEM, as well as being a member of the Science department. The students I chose to conduct my Action Research project with were a mixed ability Year 7 class that I teach for 4 Science lessons per week. I conducted the research with 18 students, 7 girls and 11 boys aged 11-12.

The 'problem'

When conducting scientific enquiry projects, sometimes referred to as APP, students tend focus on the practical rather than the planning and improving stages. They want to skip straight to 'the fun bit' of carrying out the practical, getting the writing stages over and done with as quickly and briefly as possible. If I focus on Engineering Habits of Mind will it improve their planning and evaluating grades as these tend to be where their levels are lower than other strands of assessment?

Review of current practice and literature

I have been teaching Science for 8 years and throughout that time have implemented a range of methods for both teaching and assessing students' practical and enquiry skills. Before conducting the action research I would give students a writing frame or prompts so they were aware of what they needed to include in their write up. Over the past 3 years I have focused more on making the planning and write up of Key Stage 3 'enquiry' fit the methods of assessment at Key Stage 4, such as ISAs and Controlled Assessment, so the transition from KS3-4 was a smooth one. Emphasis however has been on the planning stage, and then the conclusion and evaluation, rather than taking the time to improve whilst conducting the practical. Students too often rush their planning as they want to jump straight into the practical, and even if they conduct preliminary investigations, do not want to spend the time tweaking their method to ensure that their practical does actually work. I have previously found that despite giving the students the time to plan thoroughly, carry out a preliminary test, and then allowing them the chance to adapt their method, they will still complete the whole practical even if it doesn't work. Their evaluations of their results and method are frequently limited to quite crude descriptions of what went well and what they could do to improve it, such as ‘work better as a team’ or ‘do it more times’.

I wanted to see if spending more time, and modelling good practice, on the adapting and improving habits of mind during the preliminary testing and in the evaluation of the results and method sections, the students would improve their levels in the ‘evaluation’ strand of assessment of enquiry.

My Research Question

If I model how to evaluate as part of Year 7 projects will their “improving and adapting” get better?

The Project

The aim of my research was to compare the levels for the ‘evaluation’ strand of scientific enquiry projects taught at Key Stage 3 before and after intervention. I chose to carry out the research with a class that I was the sole teacher of. They were a mixed ability Year 7 Science class who had experience throughout the year with various enquiry projects so they were familiar with the basic format. I surveyed the students before we started the task using the student self-reports provided.

The project I chose was ‘Moon Craters’ where the students plan an experiment to look at why moon craters are different sizes. Rather than directing the students to immediately choose the one variable they wanted to change, I allowed the students to choose four different variables that they could possibly change. From that they wrote brief outlines of how they would conduct their experiments, to get them to think about the logistics of whether the experiment was possible or not. Too often students will come up with an idea for a variable but not actually think about whether it will work to answer the question they have set themselves. For example some of the students decided that they could investigate the angle that the meteor strikes the surface to see what effect it had on crater size. Although this was a legitimate variable for affecting crater size it would be very difficult to conduct, and also to make it a ‘fair test’ as it would be difficult to control speed and force. After allowing the students to choose four variables they were then tasked with eliminating two of the variables based on whether they would be able to actually conduct a ‘fair test’ experiment. The point of this was to make them really think about whether the method was achievable or if the experiment would elicit the results they were after.

Students then carried out a preliminary investigation of the two remaining variables. From this they eliminated another variable to leave them with one variable that they would use to conduct a full investigation. I allowed them time to evaluate the preliminary tests and adapt their methods to ensure they would get a good set of data to analyse. Previously they had not had the opportunity to do this as usually only 4-6 lessons are allowed for these enquiry projects. In total this time the students had 10 lessons.

Once they had adapted their methods and conducted their full investigations students were afforded the usual amount of time to draw graphs, and write a conclusion about what their results showed.

After this I spent a lesson putting the intervention in place. I showed them some example evaluations that were from generic scientific experiments, to allow them to see a variety of different ways of writing evaluations. I gave them a 'model' good evaluation that I had written that was related to the enquiry they were conducting. They spent some time discussing what made it a 'good' evaluation and how it could further be improved. They were also provided with a bullet point checklist of what they needed to include, to ensure that they had enough detail to make it a 'good' evaluation of both the method and results.

Students spent another lesson writing their evaluations, before I checked through and further gave them suggestions for improvements.

Findings

The self-report data was of mixed use. Students complete the survey both before and after the moon craters enquiry project.

The important questions for the project I was doing were;

Question 3 – I'll check and check again until I'm happy
Question 7 – I like making what I've done better
Question 12 – I work hard and practise to get better, even when it's tricky

See below for the student responses for these questions.

	Number of student responses that changed after the second survey		
Question	Up	Down	Stayed the same
3	5	3	10
7	4	3	11
12	4	10	4

From the results you can see that more students improved their scores after taking the survey for the second time for questions 3 and 7; however students' perception for question 12 actually went down after the second survey.

What was more useful was looking at the evaluation levels the students achieved after the intervention when comparing them to a previous enquiry task. Although the task is different the students still need to discuss the same areas in their evaluation. However after the intervention the majority of students improved.

One student stayed the same.
Seven students improved by 1 whole level.
Seven students improved by 2 whole levels.
Three students improved by 3 whole levels.

See appendices for further data.

Overall I believe that by showing the students what makes a good evaluation their levels in the evaluation strand of assessment have improved, some significantly.

Before I shared the levels with the students I surveyed them to see if they perceived that they had improved or not. They were asked the question "Do you think that you are now better at writing a scientific evaluation after being shown an example of a 'good evaluation' and having it explained to you? Explain your answer."

17 students believed they were better at writing evaluations. Only one gave a negative response. Here are some of the responses.

"Yes, because I didn't know what a good evaluation looked like before" – Boy 1

"Yes because I saw what an ideal answer looked like and knew how to present mine" – Girl 2

"Yes because we know what standard we want to aim for and to achieve those levels we can memorise what the guides say so we can use it in future evaluations. Additionally it helps us understand what makes a good evaluation." - Girl 5

"I think I am now better at writing scientific evaluations because I took some ideas from the good example and put it in my work" – Boy 9

"Yes because it was easier to understand and I knew what to do and it was put into paragraphs" – Girl 7

"Yes because it helped me when I looked at the one Miss Davison showed us now I'm more confident" – Boy 10

Lessons Learned and next steps

If I were to conduct this experiment again I would definitely do a few things differently. In changing the format of how to plan an enquiry project some students got very confused. As stated previously, students are normally encouraged to focus on one variable that they will be changing. By asking them to look at four variables, and then narrow it down to two, and then finally one variable, some of the students lost focus on the fact that they were only supposed to ultimately change only one variable in their experiment. I discovered some of the groups were changing both drop height and ball size in their final experiment, or drop height and surface. This meant that they were not conducting a ‘fair test’ experiment. Some students also hadn’t explained which variables they were excluding so hadn’t thought about whether their experiment would actually work. For example one group had decided to investigate force of throw, which I assumed they would eliminate after their preliminary test as there was no way they could measure the force of their throw. Another group hadn’t adapted their method after the preliminary test so their results were too close to show a distinct pattern. If I were to teach the same principals again I would build in some more time for me to actually read through their work in more detail or ask more in depth questions during feedback of ideas as I hadn’t picked up on all the confusion due to time constraints.

I would like to further expand on developing habits of mind in my teaching, especially in scientific enquiry projects, but if I were to do it again I would ensure that I do it from the start of the academic year and not part way through. This way I would not be confusing students with new ways of working just when they had become confident in how to work with the existing methods.

I would also like to build in some of the engineering habits of mind into the STEM schemes of work. STEM is taught to students in a double lesson every week in years 7-9. As they are not constrained by the National Curriculum and formal assessment teachers will have much more opportunity and freedom to explore different habits of mind in the lessons without worrying about ensuring students have covered the required content for the exams.

Appendix 1

Evaluating

Finish your plans

Look carefully at what you have planned to do. Can you actually carry out the practical and get a good set of data?

Narrow down your ideas to 2 that you could test.

Carry out a preliminary test for each of the variables.

Which variables are you going to test?
Why?

In your books you should have:

- Your four original variables that you could have investigated
- An explanation next to the 2 that you were going to exclude, giving reasons e.g. we can't do force as we can't measure the amount of force used.
- Details of how to test the two variables you had left.
  - Independent – Drop height
  - Dependent – Crater diameter
  - Control – surface, ball size etc.
- Preliminary data for those two variables – highest, lowest
- An explanation of which one you are going to exclude and why, e.g. size of ball didn't have a big enough effect on crater size
- A detailed plan for your final experiment, detailing any changes you have made to your method after your preliminary data. (evaluate your preliminary test)
- A table of results including 3 repeats

What makes a good evaluation?

- How could you have improved the way you did your practical? (Level 3)
- What improvements would you make to your method if you did it again? Why? (Level 5)
- Did the data you collected give you enough information? Why? (Level 6)
- What could you do to improve the repeatability of your data? (Level 7)

What Makes A Good Evaluation?

The results of the investigation were accurate as all my points were very close to the line of best fit, and it was possible to draw the line within all my range bars. It is possible to draw a valid conclusion from my results. My repeats were very close to each other, except at 40cm, when I had a result that was much lower than expected. I ignored this result and calculated the average without it. This means in general my results are repeatable. I checked my results with another group. We did not use the same values for the length, but our trends were both similar to each other. My trend also matches the one I found in my research. This makes my investigation reproducible.

I managed the risks fine as no one in the group got injured.
I found that my results were OK, we got a pattern and the repeats were close to each other.
My method was good. I chose the right strength of acid, but I did change the temperatures as the water baths in the room were differ temperature to those put in my plan. To improve the experiment I would use a data logger.

The method used in the experiment was the best available. We trialled the experiment first to check the values for the acid concentration were going to give a big enough spread of results. We used a measuring cylinder with a high resolution to get the best results. When timing it was hard to be accurate due to our reaction time.

My experiment was accurate. I repeated my results and found them to be precise. I checked my results with another group who had done the same experiment and my results are reproducible. I used the highest resolution possible.
I worked safely.
My method was good and no improvements are necessary.

Evaluating your practical

Level 4

- can suggest improvements to their work, giving reasons

Level 5

- makes practical suggestions about how their working methods can be improved

To move pupils from level 4 to level 5

- 1. Explain accuracy, repeatability and anomalous results.
- 2. Ensure that you can successfully identify odd-looking results (anomalies).
- 3. Look for practical and/or procedural reasons for anomalies.
- 4. Look for practical suggestions to improve procedure.
- 5. Decide whether the anomaly is an under- or over-estimate of what was
- expected.
- 6. Decide which suggested reasons for anomalies or suggestions for
- improvements best explain the results obtained.

Evaluation

- Was it a **fair test**? **Explain** your answer.
- Was it **repeatable, precise, accurate**? – How could you improve these?
- Did you make any **errors**? Were there **anomalies**?
- **How** could you improve it?
- What would you do to **extend** your investigation?

Evaluation

- I believe my test was fair because I only changed the drop height. I made sure that I used the same marble each time and the same person dropped it from the same position.
- I believe that for some results my test was precise as the crater size for drop height 20 cm was 3.5, 3.6 and 3.5. These values are all very close to each other. However the results for 40 cm weren't precise as they were 4.5, 4.6 and 3.2cm.
- If I were to repeat the experiment I would have repeated the 3.2cm value. We may have got this anomaly because the marble made the crater too close to the edge of the bowl so the sand didn't go as far.
- Our test was repeatable as we repeated it 3 times for each test height and got similar results (apart from one anomaly).
- To improve our repeatability I would repeat it 5 times and compare my results with other groups who did the same test to see if ours was similar. I would also repeat any anomalous results or exclude them form my mean.
- I think my results have answered the question as I can confidently state which drop height causes different crater sizes.
- If I were to repeat the experiment I would make the following improvements:
  - Bigger bowl so that the crater wasn't affected by it being too close to the edge
  - Keep the ruler straight in the sand using a protractor as it kept going wonky and this may have affected our results as we could have dropped it from different height causing anomalies.
  - I would use calipers to measure the diameter of the crater as the ruler kept moving the sand and this could have affected our measurements.
  - I would use a straight edge, like a ruler, to make sure I line the marble up with the correct place on the meter ruler as it was difficult to judge by eye and this may have caused uncertainties.

Appendix 2

	Question																										
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After			
Name	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	Up	Down	Same
Boy 1	2	2	3	4	2	3	4	2	3	3	4	4	4	3	3	3	4	3	4	4	3	3	4	3	2	4	6
Girl 1	2	3	2	3	3	2	2	2	1	3	3	4	2	2	2	1	2	2	4	4	2	3	4	2	5	3	4
Boy 2	3	4	3	4	2	2	2	4	2	1	3	2	3	3	2	3	3	3	2	3	2	3	2	3	7	2	3
Boy 3	2	2	2	1	2	2	1	3	2	2	3	3	2	2	1	1	2	3	1	2	2	3	2	2	4	1	7
Girl 2	2	2	3	2	2	2	4	3	2	3	3	4	1	2	1	1	4	3	4	4	3	4	2	3	5	3	4
Girl 3	2	2	3	3	2	2	2	2	3	2	2	1	3	3	2	2	2	3	4	4	2	3	3	4	3	2	7
Boy 4	2	3	3	2	2	3	2	2	2	2	3	3	2	3	2	2	3	3	2	2	3	3	4	3	3	2	7
Boy 5	2	3	4	4	1	3	2	2	3	4	4	4	3	3	2	2	4	4	3	4	3	3	2	3	5	0	7
Boy 6	3	2	2	3	1	2	2	4	1	1	2	1	2	3	2	1	4	3	1	3	3	4	4	3	6	5	1
Boy 7	4	4	3	3	4	4	1	2	2	2	4	4	3	4	2	2	3	3	3	4	2	3	4	3	4	1	7
Boy 8	3	3	4	4	2	1	2	4	4	4	3	4	2	2	2	1	3	4	3	2	3	3	4	1	3	4	5
Girl 4	2	2	2	1	2	2	1	1	1	1	3	2	3	2	2	2	3	1	4	3	2	2	3	2	0	6	6
Girl 5	3	3	3	2	2	2	2	2	4	3	4	4	3	3	2	2	2	3	3	4	2	2	3	3	2	1	9
Girl 6	2	3	3	4	2	2	1	1	3	3	3	2	2	2	1	1	2	2	2	2	3	3	4	4	2	1	9
Boy 9	2	2	2	3	3	1	3	3	2	3	1	1	3	3	1	2	3	2	4	4	2	4	2	2	4	2	6
Girl 7	2	2	4	3	2	2	2	1	1	2	3	4	3	3	2	2	3	2	4	4	3	3	4	2	2	4	6
Boy 10	2	2	3	3	2	3	1	2	1	1	2	3	3	3	2	4	2	2	3	4	2	2	4	3	5	1	6
Boy 11	3	3	3	2	2	2	3	3	4	3	2	2	4	2	3	2	3	3	2	3	4	4	3	2	1	5	6

Appendix 3

Student	First level	Final level	Change
Boy 1	3	4	1
Girl 1	3	5	2
Boy 2	3	5	2
Boy 3	3	6	3
Girl 2	3	4	1
Girl 3	4	6	2
Boy 4	3	5	2
Boy 5	4	6	2
Boy 6	4	5	1
Boy 7	4	5	1
Boy 8	3	4	1
Girl 4	3	4	1
Girl 5	5	5	0
Girl 6	3	6	3
Boy 9	4	6	2
Girl 7	4	6	2
Boy 10	3	4	1
Boy 11	3	6	3

Appendix 4

2nd June 2015 Evaluations

Practical Evaluation

improvements

How?

The inprofments we need to make are the numbers need to be even and equal amount of - counts. Our method can be inproved by not making too many mistakes on our drop test and to make it more simple than making it so difficult. Our repatability was quite wonky and not so accurate because we were measuring the crator and it was all different after each test. We could have anomilies especally when we cannot work as a team when they are bouncing the balls and not concentrating enough.

Results Evaluation

The accuary of measurments is the equipment used like mm on a ruler, sand, stander, type of ball(s) and your table. We also for our pratical - could of added a tube so the balls don't go flying everywhere. So if we could have a change of practical we could've done that! If we could have a bigger bowl of sand then - we could have a tennis ball and tube then

Tuesday 2nd June 2015

Evaluation

I think if I had to make my work better I would get a bigger bowl of sand so it would be more accurate when measuring because if you had a small bowl the measuring stick would get in the way of the crater. We could of done more than 3 tests to increase the accuracy. Our test is accurate, repeatable but we did have an anomalous result which was the 100cm drop height because it didn't continue the pattern like it was ment to instead it went down 7cm.

How else could you improve your method to get better results?

The problem of our experiment is that the ruler is hard to measure the crater and the ball went all over the place because it wasn't very sterdy.

We could have prevented it by getting a tube and putting the marble down there and to prevent the ball going everywhere.

We got these results because it wasn't very

Conclusion

our
I think are test went well but,
the things we could change
are the accuracy from where
we dropped it from because we
didnt do it accurately. another thing
we could change is the height
and make them go up in 50, 100, 150,
200. because it will be easier to
to find where it is on the
meter stick. a good thing is the
ball size we choose because
you can tell the differances
between when we drop it from
and the quater it makes.

How could you improve the accuracy?

Appendix 5

Do you think that you are now better at writing a scientific evaluation after being shown an example of a "good evaluation" and having it explained to you? Explain your answer.

Yes, because we know what standard we want to aim for and to achieve those levels we can memorize what the guides say so we can use it in future evaluations. Additionally it helps us understand what makes a good evaluation.

Do you think that you are now better at writing a scientific evaluation after being shown an example of a "good evaluation" and having it explained to you? Explain your answer.

yes because I saw what an ideal aunser looked liked and knew how to present mine.

Do you think that you are now better at writing a scientific evaluation after being shown an example of a "good evaluation" and having it explained to you? Explain your answer.

Yes because it was easier to understand and I new what to do and it was put into Paragraphs.