

## Activity L

# Quality evaluations?

Rod Watson  
King's College, University of London

## Introduction

Pupils can learn how to evaluate by simply doing them. A much more focused way of learning how to evaluate is to make explicit the qualities of good evaluations so that pupils acquire criteria for judging quality. Pupils may be concerned about their feelings about the enquiry or how well they have worked with others in their group, rather than judging an enquiry by the quality of evidence collected and how that evidence has been used. In this activity, pupils observe a simple demonstration and then look at the evaluations from ten different groups and highlight good and poor points of evaluation. They then take part in a whole class discussion focused on the quality of evaluations.

## Objectives

Pupils will learn about the features of a good evaluation, and what should be omitted from an evaluation.

## Outcomes

By the end of the lesson, pupils will be able to:

- write evaluations that highlight whether adequate controls have been made in a fair-test enquiry;
- identify whether measurements are accurate and how accuracy can be improved;
- recognise that statements of opinion about how they felt about an enquiry have nothing to do with the quality of evidence collected.

## Notes for Teachers

The following books provide a useful introduction to how to teach the concepts of evidence introduced in these materials:

- Watson, R. and Wood-Robinson, V. (2002) *Investigations: Evaluating evidence* pp. 89-107 in (eds.) Sang, D. and Wood-Robinson, V. *Teaching Secondary Scientific Enquiry*. London, John Murray.

- Goldsworthy, A., Watson, J. R. and Wood-Robinson, V. (2000) *Investigations: Developing Understanding* Hatfield, Association for Science Education, pp. 50-55 and pp. 65-68.

Another example of an exercise to teach evaluation of fair tests can be found in:

- Goldsworthy, A., Watson, J. R. and Wood-Robinson, V. (2000) *Investigations: Developing Understanding Unit 20b: Evaluation of a fair test*, pp. 113-114. Hatfield, Association for Science Education.

## Teaching Sequence

- *Task 1 (10 minutes)*

First demonstrate the candle experiment. The question being investigated is ‘Do different coloured candles burn for different lengths of time?’ The purpose of the demonstration is to set a context for looking at the quality of evaluations.

Fix two different coloured birthday candles to a heat-proof mat by standing each in a drop of melted wax. It is best to use half candles otherwise the demonstration takes too long. Light the candles and time how long it takes for each to burn down completely. The times will probably be very similar. Discuss with the class how you would know whether the difference in burning times was significant or just due to a draft or slightly different sized candles etc.

Emphasise that the class will look at the quality of evaluations and that good evaluations focus on the quality of evidence and how it has been used. Discuss the sorts of things that could be done to improve the quality of evidence e.g. better control of variables, improving accuracy by taking repeat readings and improving accuracy by better measurement. Ask pupils not only to identify what to improve, but also how to improve it. The purpose of this discussion is to focus on what goes in a good evaluation.

- *Task 2 (20 minutes)*

Hand out pupil activity sheet 1, *Quality Evaluations?* and ask the pupils to complete the sheet in pairs. The answers can be found in appendix 1. Blue and yellow highlighters have been used on the answer sheet because these colours are distinguishable on a black and white printed copy.

- *Task 3 (10 minutes)* Discuss answers with the whole class. Make the points below.

Evaluation is **not** about:

- whether the enquiry is easy or difficult; (2),
- getting the right answer and who is to blame for any errors; (5, 6, 7),
- working well together; (7),
- whether predictions were correct. (8).

Evaluation is about quality of evidence:

- Experimental design, in this case 'fair testing'. Fair testing requires relevant control variables to be identified and controlled (1, 8) and saying how they can be controlled (3, 9).
- Accuracy of data collection (5) and how accuracy can be achieved by collecting repeat readings (3, 7) or making measurements accurate (9).

Evaluations that identify how the quality of evidence can be improved are better than those that simply identify possible problems with evidence.

[**NB** Numbers in brackets refer to the statements on pupil activity sheet 2]

## Quality evaluations?

When you write an evaluation of an enquiry what should you write? What is the difference between a good quality evaluation and a poor one?

Groups of students were asked to find out whether different coloured birthday candles burn for different lengths of time. They were then asked to describe their enquiries and write an evaluation of their enquiries.

Your job in this assignment is to look at their evaluations and decide which ones are the best and explain why.

On sheet 2 there are statements from the evaluations that the students wrote. Some statements show that the student has thought carefully about the evidence. Other statements are not critical of the evidence, but give a personal opinion without any evidence to back them up.

### What you have to do:

1. Read each statement on sheet 2.
2. Work in pairs to decide which parts are *good* points of evaluation.  
Underline these good points in **red**.
3. Work in pairs to decide which parts are *poor* points of evaluation.  
Underline these poor points in **blue**.

## Statements from Evaluations

1. We tested a blue and a red candle and measured the time it took for them to burn. We tried to keep everything the same, except for the colour of the candle, but I lit the red one and Jemma lit the blue one. We should have had the same person lighting both candles.
2. I think this experiment was really easy. My results and Sue's together gave us a good pattern. We didn't include Dan's because his were way out. He must have read the stop-clock wrong all the time.
3. We tested three yellow and three pink candles. We measured the length of each candle to make sure they were all the same length. The results with different candles were different.

	Time to burn yellow candle (s)	Time to burn white candle (s)
First candle	183.03	185.95
Second candle	242.11	181.73
Third candle	187.56	189.44
Average	204.23	185.71

4. We think the candles were different thicknesses and we should have weighed them to make sure they were the same size.
5. We tested one blue and one white candle and timed them to see how long they burnt for. We repeated the measurements three times with new blue and white candles to make it fair.

Pupil activity sheet 2: Quality evaluations?

6. It was the right answer, because we did it fairly. We made sure the candles were all the same size and our timing was accurate. It wasn't our fault that people kept walking by and making drafts, so you can't blame us and say we weren't fair.
7. Our experiment was good because it proved something. We now know that red candles burn faster than blue ones.
8. I feel very pleased with the success of the experiment. We worked together well as a group and did the investigation very well. I would have liked to come to a much more final conclusion before having to finish. This is not our fault. We did manage our time well and got a lot done in the time we had.
9. We could have investigated further by timing more candles and taking an average, but I think we have the correct results and carried out a well planned investigation, which was worthwhile and interesting.
10. I think I have done well in predicting the results. I said the red candles would burn longer and they did. We made it as fair but we couldn't tell if the wicks were exactly the same.
11. I used a stopwatch to make the experiment accurate. I also measured the length of each candle with a ruler and they were the same length (4.3cm). I trimmed the wicks of the candle before I started to make sure they were all the same length. The flames of the candles kept on blowing around because Joel kept opening the window. This mucked up our experiment and made the wax run. If I repeated the experiment I would shield the candle with books.
12. I don't think this tells me much. The yellow candle burnt for 240.45secs and the blue one for 242.11secs. This might have been because the candles were just a bit different and had nothing to do with the colour of the candle. I need to do it again with a lot more candles to find out if there is really a difference.

## Appendix 1: Chemical Reactions & Measurement

### Answers to pupil activity

Answers are coded as follows:

Yellow - good points of evaluation.

Blue - poor points of evaluation.

1. We tested a blue and a red candle and measured the time it took for them to burn. We tried to keep everything the same, except for the colour of the candle, but I lit the red one and Jemma lit the blue one. We should have had the same person lighting both candles.

*Evaluation comments are related to fair testing, but the person who takes the measurement is not relevant.*

2. I think this experiment was really easy. Sue's and my results together got a good pattern. We didn't include Dan's because his were way out. He must have read the stop-clock wrong all the time.

*Opinions about whether the experiment is easy have nothing to do with the quality of the evidence. No evidence is given to be able to judge whether different students had made measurement errors. Dan may have been 'way out', but out voting another is not a good way of evaluating.*

3. We thought that the candles might not be exactly the same so we tested three yellow and three pink candles and took the average to get a more accurate result. We measured the length of each candle to make sure they were all the same length. The results with different candles were different:

	Time to burn yellow candle (s)	Time to burn white candle (s)
First candle	183.03	185.95
Second candle	242.11	181.73
Third candle	187.56	189.44
Average	204.23	185.71

We think the candles were different thicknesses and we should have weighed them to make sure they were the same size.

*Procedures for increasing accuracy by averaging are mentioned, as are two factors affecting whether relevant variables were controlled.*

4. We tested one blue and one white candle and timed them to see how long they burnt for. We repeated the measurements three times with new blue and white candles to make it fair.

*The writer has identified the need to make repeat readings, but justifies this in terms of fair testing (i.e. controlling variables) rather than in terms of accuracy.*

## Appendix 1: Chemical Reactions & Measurement

5. It was the right answer, because we did it fairly. We made sure the candles were all the same size and our timing was accurate. It wasn't our fault that people kept walking by and making drafts, so you can't blame us and say we weren't fair.

*This statement contains comments on how well the group had worked, which has nothing to do with quality of evidence. Fair testing and accuracy are considered.*

6. Our experiment was good because it proved something. We now know that red candles burn faster than blue ones.

*This statement has nothing to do with quality of evidence.*

7. I feel very pleased with the success of the experiment. We worked together well as a group and did the investigation very well. I would have liked to come to a much more final conclusion before having to finish. This is not our fault. We did manage our time well and got a lot done in the time we had. We could have investigated further by timing more candles and taking an average but I think we have the correct results and carried out a well planned investigation, which was worthwhile and interesting.

*These students are trying to justify how they worked, but do make a suggestion for improving the investigation.*

8. I think I have done well in predicting the results. I said the red candles would burn longer and they did. We made it as fair but we couldn't tell if the wicks were exactly the same.

*This student judges the enquiry in terms of whether her prediction was correct rather than in terms of the quality of the evidence. One uncontrolled variable is mentioned in relation to fair testing.*

9. I used a stopwatch to make the experiment accurate. I also measured the length of each candle with a ruler and they were the same length (4.3cm). I trimmed the wicks of the candle before I started to make sure they were all the same length. The flames of the candles kept on blowing around because Joel kept opening the window. This mucked up our experiment and made the wax run. If I repeated the experiment I would shield the candle with books.

*The quality of the evidence is considered in terms of accuracy of measurement.*

10. I don't think this tells me much. The yellow candle burnt for 240.45secs and the blue one for 242.11secs. This might have been because the candles were just a bit different and be nothing to do with the colour of the candle. I need to do it again with a lot more candles to find out if there is really a difference.

*Uncontrolled variables are mentioned, although no specific ones are identified.*