ISEB
Independent Schools Examinations Board

CE SCIENCE

THINKING AND WORKING AS A SCIENTIST: MODELLING SCIENTIFIC ATTITUDES

Pupils should:

- be given opportunities to think about objectivity when measuring, with increasing awareness of the needs for accuracy, precision, repeatability and reproducibility; they should consider the validity of experimental results in terms of fair testing

Thinking and Working as a Scientist and the Key Stage 3 programme of study for the National Curriculum introduce a number of new terms, that replace the more general term 'reliability, when evaluating experiments. These terms are: accuracy, precision, repeatability and reproducibility.

This paper provides reasons for the change and the implications for teaching and assessment.

The changes have been recommended by the Association of Science Education and have the support of the government and the examination awarding bodies. Thus, Common Entrance is being brought in line with the agreed thinking of the science education community. Further information can be obtained in the book 'The Language of Measurement', obtainable from the ASE.

The word 'reliability' has two distinct everyday meanings, "consistently good in quality or performance", and "able to be trusted". Within Science, the word 'reliable' has been applied to different things: such as a set of results, patterns in data, to conclusions drawn from the data, as well as the quality of information sources. It is a convenient word, but ambiguous.

It has been decided, after consultation with experts, to replace the word 'reliable' with 'repeatable' and 'reproducible'.

Conclusions drawn from data can now be expressed in terms of the amount of confidence the pupil has in the quality of the evidence. This requires pupils to be taught to ask key questions when looking at data:

- Can I rely on the data when drawing conclusions?
- Are the uncertainties and inconsistent measurements in my data small enough not to affect the quality of my conclusions?
- Does the difference between one measurement and another, reflect a real change in the factor being measured?

At the heart of this thinking is that all measurements are subject to errors. This is rather more than simply failing to follow the instructions carefully. It is likely that the equipment we use to measure our variables, or the way in which we use the equipment, may also introduce errors into the results. Our experimental measurements are, at best, an approximation of the "true" measurements.

We need to encourage our pupils to accept that their results will always contain some errors, and to reflect on the reasons for them. Wherever possible, can pupils think of ways of reducing the errors?

The ASE define the key terms used in Thinking and Working as a Scientist. What follows uses their definitions:

ACCURACY, PRECISION

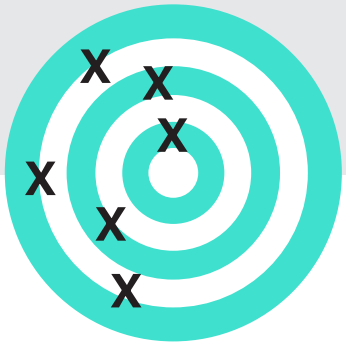
A measurement is accurate, if it is close to the true value for that measurement. Think of it in terms of the dart board below, where the true value is at the centre of the dart board.

Measurements (each individual cross) are **accurate** when they are close to the centre of the board.

High Accuracy
High Precision

Low Accuracy
High Precision

High Accuracy
Low Precision



Low Accuracy
Low Precision

The further the further the measurements are away from the centre, the lower their accuracy.

Precision defines how closely the different measurements agree with each other. If the measurements cluster closely together, then they are 'precise'.

In the second diagram, the results are precise, but with a low accuracy, which means that the results cluster together, but are not close to the true value in the centre of the board.

In the first diagram, the results are precise and accurate, which means they cluster closely together around the true value in the centre of the board.

We say that the results are **valid** if they measure what they are supposed to be measuring and are accurate and precise. This requires careful procedures and good quality measuring instruments. It will also require that their experiments are fair tests.

Fair tests have controlled variables, so that changes to the independent variable cause changes to the dependent variable.

We can estimate the accuracy of a set of results by comparing our mean value with means produced by other groups or by repeating the experiment many times.

We can measure the precision of the results by calculating the range (the maximum value minus the minimum value) of our results.

The best way to compare results between groups is to use the mean and the range.

REPEATABLY, REPRODUCIBILITY

Imagine that the same group of people carry out the same measurements, using the same apparatus, in the same conditions, time and time again. We would expect their result to be precise. If every set of results clustered together, then we would say that the group's results were **repeatable**.

It is quite possible that a group in another part of the room, or in a different class or school, are also repeating the experiment, obtaining precise results that are repeatable.

However, the means and ranges of the two groups could be **different** to each other. Perhaps one group's results are closer to the true result than the other, or perhaps neither group is close to the true result.

When two groups of people obtain similar results, then we say their results are **reproducible**. This is what scientists are looking for – groups in different laboratories producing similar, reproducible results. This gives scientists the confidence to believe that their results are close to the "true" value, and that their results have **validity**.

Summary of key terms

Accuracy: Results that are close to the 'true result"

Precision: Repeated results that are consistently close together

[Precise results may not be accurate!]

Repeatability: Repeated results that are consistently close together (precise), when obtained by the same individual or working group.

Reproducibility: Repeated results that are consistently close together (precise), when obtained by different individuals or working groups using the same or similar equipment. Results that are reproducible when using different methods, suggest that the results are close to the "true value" and are both precise and accurate.