Mathematics

Errors and uncertainties

The examples that follow are only illustrations of the depth required to address uncertainties and errors. They do not represent all the ways and means to deal with uncertainties and errors.

The significance of errors and uncertainties

Data collection and data analysis are central to the scientific process. All data are limited in the information they convey; this means that there is uncertainty in every measurement. This uncertainty is expressed as a range of possible values within which the true value lies.

Moreover, experimental design always involves some weaknesses and assumptions that may give rise to errors. These uncertainties and errors have an impact on the validity of scientific findings, and so they must be carefully communicated and considered in the evaluation of the results.

Students should, therefore, be encouraged to record uncertainties and errors, and to consider their impact on the results of **all** their experimental work during the course. The criteria for the internal assessment component of the course also include these skills.

The resources here explain the steps involved in handling errors and uncertainties, so that these processes can be integrated into students' learning. Experimental data continually expand the boundaries of science, but an awareness of the limits of that knowledge is crucial to its application.

Types and sources of error

There are two types of errors: systematic and random.

Systematic errors

The first category of errors concerns systematic errors, which:

- are errors due to identifiable causes in the experimental design
- give results that are consistently higher or lower than the true value
- are not reduced by repetition of the experiment

- can, in principle, be reduced by modifications to the experiment.

Examples of causes of systematic error include:

- error caused by poor insulation during thermochemical experiments
- error caused when measuring gas volume by collection over water, assuming the gas is insoluble.

Random errors

The second category of errors involves random errors, which:

- arise from the limit of the precision of the experimental apparatus
- lead to measurements that are equally likely to be higher or lower than the true value
- can be minimized by measurement repetition and averaging the measurements, leading to cancelling out of the variation
- can be reduced with the use of more precise measuring equipment
- can be quantified in all measurements and are expressed as a ± range of values

The following are examples of measurements showing a decrease in random error and thus an increase in precision (the term "precision" is explained in the following section).

5.2 g ± 0.1 g

5.13 g ± 0.01 g

5.312 g ± 0.001 g

↓ decreasing random error

increasing precision

Accuracy and precision

Accuracy is how close a measured value is to the correct value. Experiments with smaller systematic errors are more accurate.

Precision indicates how many significant figures there are in a measurement. Data with smaller random errors are more precise.

For example, the normal boiling point of water is 100°C. Measurements from two experiments are provided in the table on boiling point versus uncertainty.

	Boiling point (°C)	**Uncertainty**
Experiment 1	99.5°C	± 0.5°C
Experiment 2	98.15°C	± 0.05°C

Measurements from two experiments on boiling point of water vs uncertainty

Experiment 1 is more accurate but less precise. Experiment 2 is less accurate but more precise. (Note the consistency in the number of decimal places between each measurement and its uncertainty. The term "uncertainty" is explained in the next section.)

Estimating and recording uncertainty in raw data

All raw data should be given with an associated uncertainty (random error). As it is the last digit that is uncertain, the uncertainty must be expressed to the same number of decimal places as those in the measurement.

Example 1

34.0 cm_3 ± 0.5 cm_3

34.10 cm_3 ± 0.05 cm_3

It can sometimes be difficult to determine the uncertainty associated with a measurement as several factors may influence the reading. Examples include the reaction time of the experimenter when measuring time, or judging the colour change at the end point of an indicator during a titration. These may not be quantifiable but should be noted as additional sources of error.

In general, random error can be estimated as follows.

- The uncertainty for a specified temperature is often stated by the manufacturer of the instrument or glassware.
- Where not stated, for digital equipment, the uncertainty is the smallest scale division (sometimes known as "the least count").
- Where not stated, for analogue equipment, the uncertainty is half the smallest division.

Example 2

For example, consider the following data obtained from experiments to measure the mass of two samples using a digital balance. According to the table on data from the two-decimal point balance, the uncertainty is ±0.01 g.

	Mass (g ± 0.01 g)
Sample 1	1.23
Sample 2	0.95

Uncertainty in measurements when using a two-decimal point balance

However, as the table on data from the three-decimal point balance shows, the uncertainty is ±0.001 g.

	Mass (g ± 0.001 g)
Sample 1	1.233
Sample 2	0.954

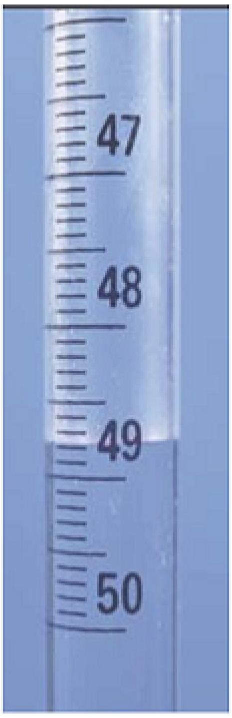
Uncertainty in measurements when using a three-decimal point balance

Note that when the same apparatus is used for a set of data, the uncertainty can be recorded in the column header, as it is the same for each reading.

Example 3

As the figure of the glassware shows, the smallest division (or least count) is 0.1 cm_3. In this case, the uncertainty is taken as ± 0.05 cm_3.

Therefore, the volume poured is expressed as 48.80 cm_3 ± 0.05 cm_3.



Consideration of uncertainties in processed data

The uncertainties associated with each measurement are known as "absolute uncertainties". When data values are processed, the impact of these uncertainties must be considered through the calculations. This process, which is sometimes referred to as "error propagation", generates a final or overall uncertainty for the experimental results. The general principle is that the overall uncertainty is the sum of the absolute uncertainties. The process used in error propagation depends on the nature of the steps in the data analysis.

Error propagation for addition or subtraction of measurements

When data values are added or subtracted, the uncertainties associated with each value must be added together. This is because the total error must include the range from the possible maximum to the possible minimum of each reading.

For example, consider the table showing data from an experiment to determine change in temperature.

	Temperature (°C ± 0.1°C)
Final temperature	29.9°C
Initial temperature	27.9°C

Experiment to determine change in temperature

According to the data,

Temperature change = (29.9°C ± 0.1°C) – (27.9 ± 0.1°C) = 2.0°C ± 0.2°C

Rationale:

Final temperature is in the range 29.8°C to 30.0°C

Initial temperature is in the range 27.8°C to 28.0°C

Therefore, the temperature difference could be as high as 2.2°C (30.0°C – 27.8°C) or as low as 1.8°C (29.8°C – 28.0°C), namely 2.0°C ± 0.2°C.

Error propagation for multiplication or division of measurements

When data values are multiplied or divided, or when measurements with different uncertainties are made, we must first standardize these absolute uncertainties before they can be combined. The following steps apply.

- Convert each absolute uncertainty into a percentage uncertainty:

$$\text{Percentage uncertainty} = \frac{\text{absolute uncertainty}}{\text{data value}} \times 100$$

- Add the percentage uncertainties for each data value.
- Express overall uncertainty as percentage uncertainty, or convert it back into a final absolute uncertainty.

For example, see the table on an experiment to determine *n*, the number of moles in a solution.

	Data value	Absolute uncertainty	Percentage uncertainty
Concentration	1.00 mol dm_{-3}	± 0.05 mol dm_{-3}	$\frac{0.05}{1.00} \times 100 = 5\%$
Volume	10.0 cm_3	± 0.1 cm_3	$\frac{0.1}{10.00} \times 100 = 1\%$

Calculating the absolute and percentage uncertainties when determining the number of moles in a solution

$$
\begin{aligned}
n(\text{mol}) &= \text{concentration (mol dm}^{-3}) \times \text{volume (dm}^{3}) \\
&= 1.00 \text{ mol dm}^{-3} \times \tfrac{10.0}{1{,}000} \text{ dm}^{3} \\
&= 0.0100 \text{ mol}
\end{aligned}
$$

$$
\begin{aligned}
\text{Total uncertainties} &= 5\% + 1\% \\
&= 6\%
\end{aligned}
$$

$$
\begin{aligned}
\text{Converting back to absolute value} &= \tfrac{6}{100} \times 0.0100 \text{ mol} \\
&= 0.000600 \text{ mol}
\end{aligned}
$$

The answer can be given with total uncertainty expressed either as a percentage or as absolute value.

$$
\begin{aligned}
\text{n(mol)} &= 0.0100 \text{ mol} \pm 6\% \\
&= 0.0100 \pm (6 \times 10^{-4}) \text{ mol}
\end{aligned}
$$

In calculations involving multiplication and division, the precision of the result is limited by the precision of the least precise measurement. So, the result should be given to the same number of significant figures as those in the least precise data value.

While considering the significant figures in the uncertainty, the common protocol is as follows.

- When the final percentage uncertainty is greater than or equal to 2%, it should be given to one significant figure.
- When the final percentage uncertainty is less than 2%, it should be given to not more than two significant figures.

For example, consider the table on an experiment to determine concentration.

	Data value	Absolute uncertainty	Percentage uncertainty
Mass of solute	6.5 g	± 0.1 g	$\frac{0.1}{6.5} \times 100 = 1.5\%$
Volume of solution	63.10 cm_3	± 0.05 cm_3	$\frac{0.05}{63.10} \times 100 = 0.0792\%$

Calculating the absolute and percentage uncertainties when determining concentration

$$\text{Concentration (g cm}^{-3}) = \frac{\text{mass (g)}}{\text{volume (cm}^3)} = \frac{6.5 \text{ g}}{63.10 \text{ cm}^3} = 0.103 \text{ g cm}^{-3}$$

$$\text{Total uncertainties} = 1.5\% + 0.0792\% = 1.5792\%$$

As the least precise data (mass) have two significant figures, the result is given to two significant figures. As the uncertainty is <2%, the uncertainty is given to two significant figures.

Therefore,

Concentration = 0.10 g cm_{-3} ± 1.6%

Error propagation for exponents *(HL only)*

Consider calculations where a data value is raised to the power of a whole number. If this whole number is *n*, a typical calculation might be represented as follows.

Result = (data value) $_n$

Note that the units associated with the data value are also raised to the power *n*.

To propagate the uncertainty in this type of calculation, it is best to first convert the absolute uncertainty into a fractional (relative) uncertainty.

$$\text{Fractional uncertainty} = \frac{\text{absolute uncertainty}}{\text{data value}}$$

This fractional uncertainty is then multiplied by the exponent *n* to give the total uncertainty.

(Rationale: This is really a form of multiplication and so, again, the process involves adding the relative uncertainties—in these cases, by the same number of times as the exponent.)

For example, consider the table on a rate experiment to investigate the order of the reaction with respect to reactant A.

[A] (mol dm$_{-3}$)	**Absolute uncertainty (mol dm$_{-3}$)**	**Fractional uncertainty**
0.30	± 0.02	$\frac{0.02}{0.30} = 0.0667$

Absolute and fractional uncertainties of a rate experiment to determine the order of the reaction with respect to reactant A

The calculation involves raising the concentration of reactant A to the power 3 (i.e. cubing it). The fractional uncertainty in the concentration of A must then be multiplied by 3 to arrive at the total uncertainty.

$$[\text{A}]^3 = (0.30)^3(\text{mol dm}^{-3})^3 \quad \text{Total uncertainty} = 0.0667 \times 3$$
$$= 0.027 \text{ mol}^3 \text{ dm}^{-9} \qquad\qquad = 0.20$$

Note that 0.20 is a fractional quantity, and so it has no units.

Result = 0.027 mol$_3$ dm$_{-9}$ ± 0.2

The total uncertainty can be converted back to an absolute uncertainty, as shown below.

0.20 × 0.027 mol$_3$ dm$_{-9}$ = 0.0054 mol$_3$ dm$_{-9}$

Alternatively, it can be converted to a percentage uncertainty.

$0.20 \times 100 = 20\%$

Final result = 0.027#±#0.005 mol_3 dm_{-9} or 0.027 mol_3 dm_{-9} ± $(2 \times 10_1)\%$

If a calculation involves extracting the *n*-th root of a value, the process is similar, but the fractional uncertainty is now divided by *n*.

Error consideration when taking averages of data values

Repeated measurements can be used to determine an average value for a result. In this case, the final uncertainty is the same as the uncertainty in the component values.

For example, consider the table on calculation of Δ*H* in an experiment. In this case, three data values are generated, all with the same uncertainty.

Observation 1 Δ*H* (kJ mol_{-1})	+100 kJ mol_{-1} ± 10%
Observation 2 Δ*H* (kJ mol_{-1})	+110 kJ mol_{-1} ± 10%
Observation 3 Δ*H* (kJ mol_{-1})	+108 kJ mol_{-1} ± 10%
Average Δ *H*	**+ 106 kJ mol $_{-1}$ ± 10%**

Error consideration when using average ΔH

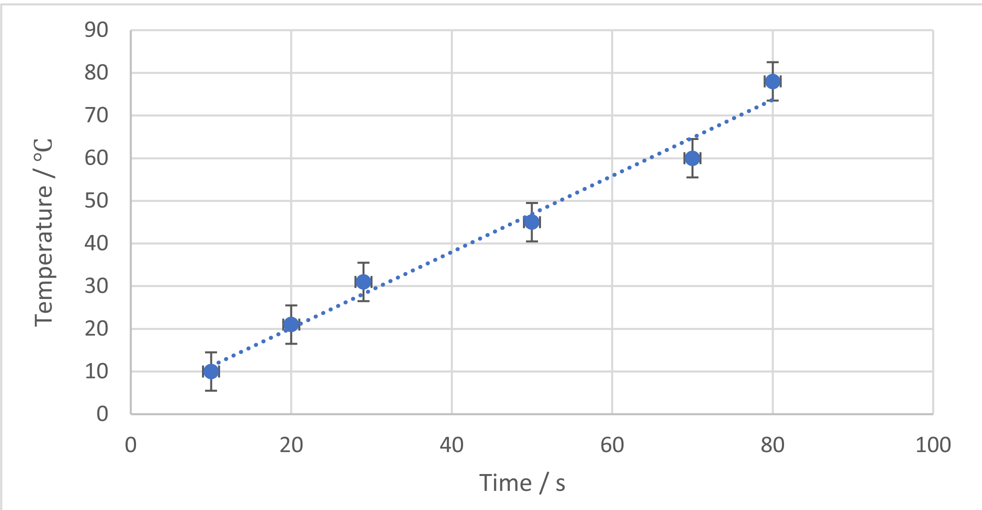
The final uncertainty is the same as the uncertainty of each data value.

(Rationale: The uncertainty is not increased by taking repeat measurements; in fact, the experiment is repeated to reduce the random error.)

Graphical representation of uncertainties

Uncertainty bars

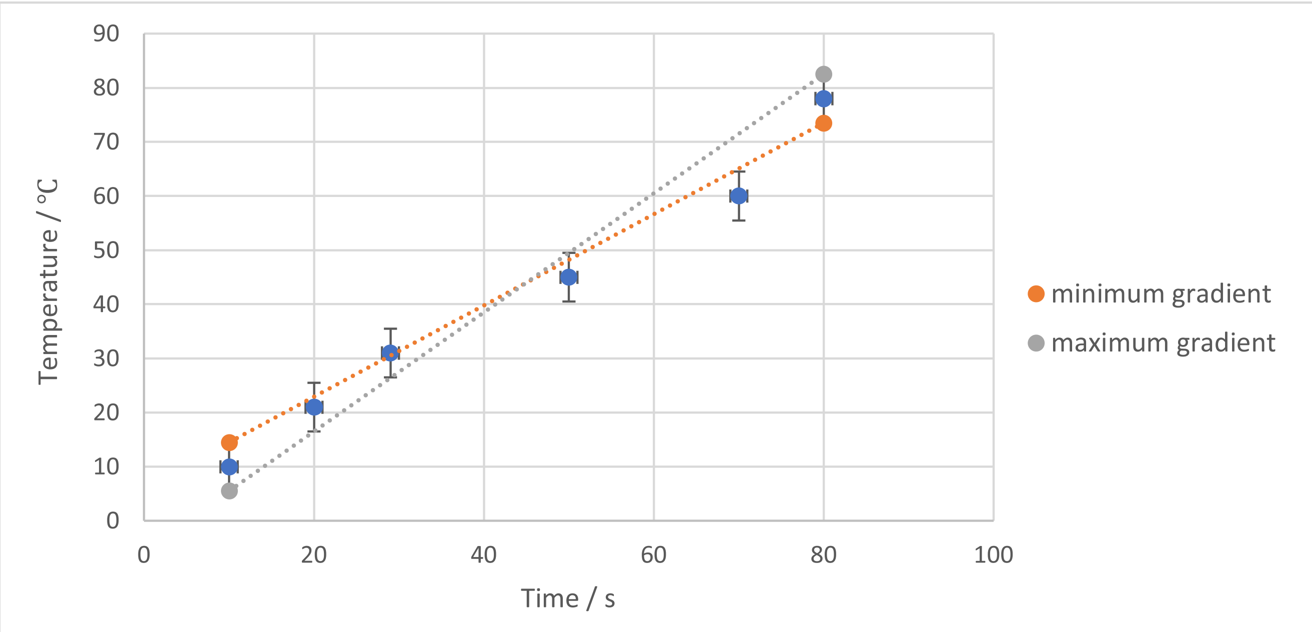
An effective way to communicate the uncertainties associated with data values is to use uncertainty bars (also called “error bars”) in graphs. These bars join together points for the maximum and minimum of the range covered by the uncertainty, above and below the data value. They can be used on the axes of either or both the dependent (*y*) variable and the independent (*x*) variable. Uncertainty bars can be plotted manually from the data values and uncertainties. They are generally plotted using graphing software. For example, consider an experiment on change in temperature over time during a reaction. As the figure on the uncertainty bars for this experiment shows, the uncertainty of each temperature data point (displayed by the uncertainty bars) is ±#5°C.



This example also shows uncertainty bars for the data values of time on the *x*-axis, but these are obviously very small compared to the uncertainties in temperature.

It is often the case that the uncertainty in one measurement is much greater than that in the other. In such cases, the total uncertainty can be taken as being due to the larger uncertainty alone, as the smaller value will have negligible impact.

Uncertainty bars can be used to calculate the gradient of the best-fit line for the given data points. As the figure on gradient of the best-fit line shows, this is done by drawing two lines: one with the minimum gradient and the other with the maximum gradient. Both lines go through the uncertainty bars. (For obvious reasons, these are sometimes called “lines of worst fit”.) Graphing software can be effective here too, or the lines can be drawn by hand.



The value of the final gradient of the best-fit line is the average of the gradients of the maximum and minimum lines.

$$\text{Final gradient} = \frac{\text{gradient of maximum line} + \text{gradient of minimum line}}{2}$$

$$\text{Final gradient uncertainty} = \frac{1}{2}(\text{maximum gradient - minimum gradient})$$

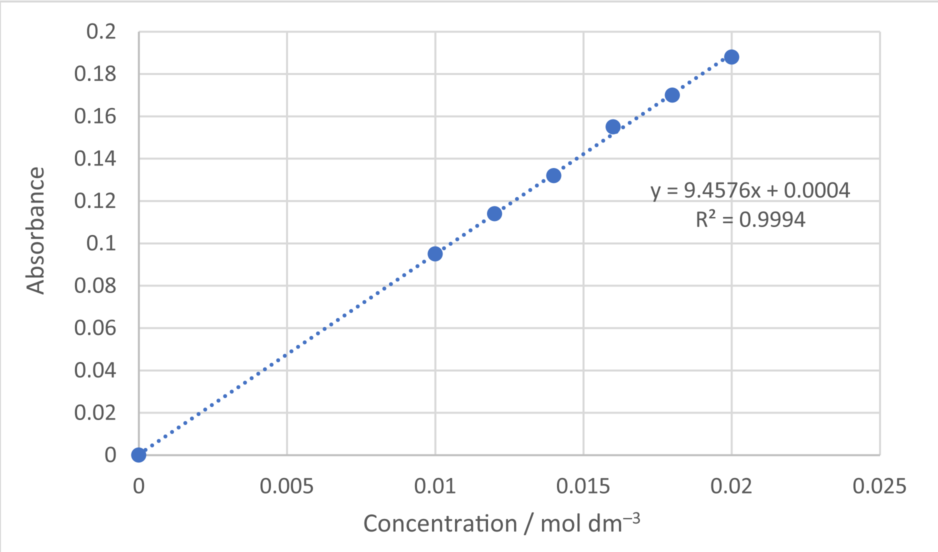
Coefficient of determination (*R* ₂)

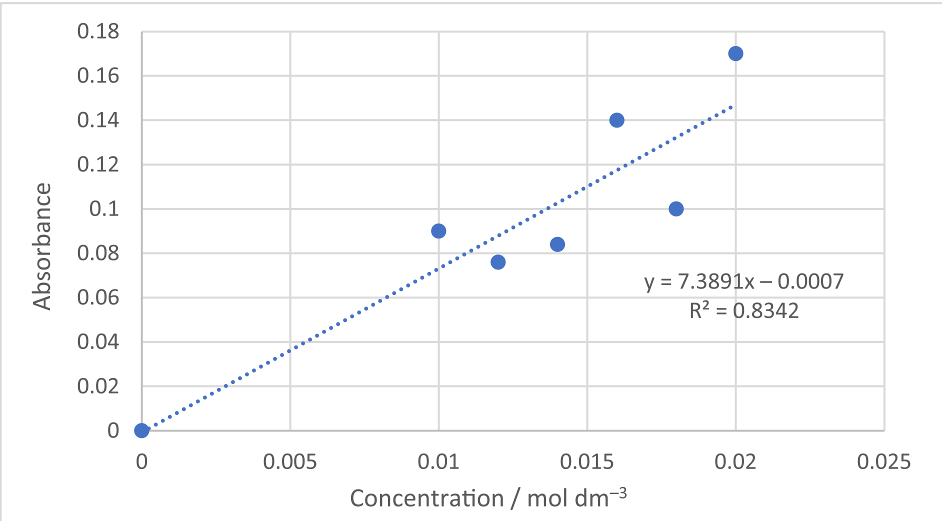
A best-fit line or best-fit curve is a trendline through a scatter plot of data points that best expresses the relationship between the points. Graphing software can often be used to generate the best-fit line by calculating its gradient and intercept in the equation $y = mx + c$. The position of the line can also be estimated manually by making a judgement of its optimum gradient and intercept, taking into account data points that may lie on either side of the line. Uncertainty bars help to make this process more accurate.

The next step is to assess how well the line or curve is a true representation of the data points. This can be done using a statistical tool known as R_2, the coefficient of determination. Students are not expected to be able to calculate R_2 values, as these are easily generated by graphing software. The important application is in interpreting their meaning.

R_2 values range from 0.0 to 1.0. An R_2 value of 1.0 indicates that the data fit the linear or curve equation perfectly. In other words, every data point lies on the best-fit line or best-fit curve. R_2 values less than 1.0 indicate that at least some variability in the data cannot be accounted for by the model. Simply put, the higher the value, the better the fit of the line with the data points.

For example, consider the figure on the best-fit lines for concentration versus absorbance data from two experiments. In the case of the first experiment, the value of R_2 is 0.9994. The best-fit line almost passes through all the data points. Note that it is likely that this best-fit line would pass through the origin if a calibration were to be carried out. Most graphing software allows users to implement this change. For the second experiment, the value of R_2 is 0.8342. Here, the best-fit line does not pass through all the data points. In fact, the data points are scattered on either side of the line.





Error analysis and evaluation

The evaluation of the conclusion of an experiment must include consideration of both systematic errors and the total uncertainty from the propagation of errors through data analysis.

Where a literature or theoretical value exists for the experimental result, the difference between the experimental and literature values can be calculated as a percentage.

$$\text{Percentage difference} = \frac{\text{experimental value - theoretical value}}{\text{theoretical value}} \times 100$$

This difference is known as the "percentage error" or "experimental error". The higher the experimental error, the less accurate the result.

The experimental error should be compared with the total uncertainty value for the experiment. If the experimental error is outside the range of the total uncertainty, it suggests that random error alone cannot explain the experimental error and that systematic errors are also present.

For example, consider the table showing data from an experiment to measure the enthalpy of a reaction.

Experimental value	**Theoretical value**	**Experimental error**
58.5 kJ mol$_{-1}$ ± 2.5 kJ mol$_{-1}$	55.0 kJ mol$_{-1}$	$\frac{(59.5 - 55.0)\ \text{kJ mol}^{-1}}{55.0\ \text{kJ mol}^{-1}} \times 100 = 6.3$

Experimental error when measuring the enthalpy of a reaction [Source: Halfman, n.d.]

$$\text{Percentage total uncertainty in the experimental value} = \frac{2.5}{59.5} \times 100$$
$$= 4.27\%, \text{ rounded to } 4\%$$

The experimental error is, therefore, larger than the total uncertainty. This means that the difference between the experimental and theoretical values cannot be explained in terms of the random errors in the measurements alone; systematic errors must also be present.

An important aspect of experimental evaluation is to suggest sources of systematic error, to determine in which direction each error may have influenced the result and to suggest modifications to the experiment to reduce these effects.