Concepts of Evidence: the Thinking Behind the Doing (GCSE version)

We are indebted to Glen Aikenhead of the University of Toronto for his detailed comments on this version and for many of the examples we use to illustrate the ideas. December 2002.

*Please quote as:
Gott, R, Duggan, S and Roberts, R. University of Durham, http://www.dur.ac.uk/richard.gott/Evidence/cofev.htm*

The section numbers are not sequential. They are references to a more extensive list. Only those ideas which have been validated as being appropriate to GCSE are included here.

1 Fundamental ideas

0	***Introduction***	Investigations must be approached with a critical eye. What sort of link is to be established, with what level of measurement and how will opinion and data be weighed as evidence? *This first category pervades the entire scheme and sets the context in which all that follows needs to be judged.*
1	***Opinion and data***	…it is necessary to distinguish between opinion based on scientific evidence and ideas on the one hand, and opinion based on non-scientific ideas (prejudice, whim, hearsay . .) on the other. *Distinguishing between the measurable energy emitted from a mobile phone mast and the 'energy' associated with 'crystals'.*
2	***Links***	…a scientific investigation seeks to establish links (and the form of those links) between two or more variables
3	***Association and causation***	…links can be causal (change in the value of one variable CAUSES a change in another), or associative (changes in one variable and changes in another are linked to some third, and possibly unrecognised, third (or more) variable)
4	***Types of measurement***	…interval data (measurements of a continuous variable) are more powerful than ordinal data (rank ordering) which are more powerful than categoric data (a label) *Being able to say that 2 wavelengths are 670nm. and 460nm. is more useful than saying one is longer than the other or that one is red and the other blue.*
5	***Extended tasks***	…measurements, for instance, can be very complicated and constitute a task on their own, but they are only meaningful when set within the wider investigation(s) of which they will form a part *The measurement of 'the absorbancy of a papertowel' involves a 'method' which, all together, contributes to the final measure of absorbancy.*

2 Observation

0	***Introduction***	Observation of object and events can lead to informed description and the generation of questions to investigate further. *Observation is one of the key links between the 'real world' and the abstract ideas of science. Observation, in our definition, does not include 'measurement' but rather deals with the way we see objects and events through the prism of our understanding.*

1	***Observing objects***	… objects can be 'seen' differently depending on the conceptual window used to view them. *…a low profile car tyre can be seen as nothing more than that, or it can be seen as a way of increasing the stiffness of the tyre, thus giving more centripetal force with less deformation and thus improving road holding.*
2	***Observing events***	… events can similarly be seen through different conceptual windows. *…the motion of a parachute is seen differently when looked at through a framework of equal and unequal forces and their corresponding accelerations.*
3	***Using a key***	… the way in which an object can be 'seen' can be shaped by using a key *e.g. a branching key gives detailed clues as to what to 'see'. It is, then, a heavily guided concept-driven observation*
4	***Taxonomies***	… taxonomies are a means of using conceptually driven observations to set up classes of objects or organisms that exhibit similar/different characteristics or properties with a view to using the classification to solve a problem. *....organisms observed in a habitat may be classified according to their feeding characteristics (to track population changes over time for instance) or a selection of materials classified into efficient conductors identified from inefficient conductors*
5	***Observation and experiment***	… observation can be the start of an investigation, experiment of survey. *… noticing that shrimp populations vary in a stream leads to a search for a hypothesis as to why that is the case, and an investigation to test that hypothesis.*
6	***Observation and map drawing***	… technique used in biological and geological fieldwork to map a site based on conceptually driven observations that illustrate features of scientific interest *....an ecologist may construct a map of a section of a stream illustrating areas of varying stream flow rate or composition of the stream bed.*

3 Measurement

0	***Introduction***	Measurement must take into account inherent variation due to uncontrolled variables, human error and the characteristics of the instruments used. *This section lies at the very centre of our model for measurement, data and evidence and is fundamental to it.*
1	***Inherent variation***	…the measured value of any variable will never repeat unless all possible variables are controlled between measurements - circumstances which are very difficult to create *Repeated bounces of a squash ball under ostensibly identical conditions will result in varied data.*
2	***Human error***	…the measured value of any variable can be subject to human error which can be random, or systematic *In the case of the squash ball, human error could result from a shaky hand when the ball was released (random) or the bounce height could always appear higher than it really is if the observer was below the height looking up at the bounce against the rule*

4 Instruments: underlying relationships

1	***Linear relationships***	...most instruments rely on an underlying and preferably linear relationship between two variables. *e.g. A thermometer relies on the relationship between the volume of a liquid and temperature.*
2	***Non-linear relationships***	...some 'instruments', of necessity, rely on non-linear relationships. *e.g. Moving iron ammeter, pH.*
3	***Complex relationships***	...the relationship may not be straightforward and may be confounded by other factors. *e.g. The prevalence, or size, of a species of lichen is an indicator of the level of pollution but other environmental factors such as aspect, substrate, or air movement can also affect the distribution of lichen.*
4	***Multiple relationships***	...sometimes several relationships are linked together so that the measurement of a variable is indirect. *e.g. Medical diagnosis often relies on indirect multiple relationships. Also, braking distance is an indirect measure of frictional force.*

5 Instruments: calibration and error

0	***Introduction***	Instruments must be carefully calibrated to minimise the inevitable uncertainties in the readings *All instruments must be calibrated so that the underlying relationship is accurately mapped onto the scale. If the relationship is non-linear, the scale has to be calibrated more often to map that non-linearity. All instruments, no matter how well made, are subject to error. Each instrument has finite limits on, for example, its resolution and sensitivity.*
1	***End points***	...the instrument must be calibrated at the end points of the scale. *e.g. A thermometer must be calibrated at 0 °C and 100 °C.*
2	***Intervening points***	...the instrument must be calibrated at points in between to check the linearity of the underlying relationship. *e.g. A thermometer must be calibrated at a number of intervening points to check, for instance, for non-linearity due to non-uniform bore of the capillary.*
3	***Zero Errors***	...there can be a systematic shift in scale and that instruments should be checked regularly. *e.g. If the zero has been wrongly calibrated, if the instrument itself was not zeroed before use, or if there is fatigue in the mechanical components, a systematic error can occur.*
4	***Overload, limiting sensitivity / limit of detection***	...there is a maximum (full scale deflection) and a minimum quantity which can be measured reliably with a given instrument and technique. *e.g. change in mass when Mg burns in air could not be detected on scales that measure to only whole grammes.*
5	***Sensitivity****	...the sensitivity of an instrument is a measure of the amount of error inherent in the instrument itself. *e.g. An electronic voltmeter will give a reading that fluctuates slightly.*
6	***Resolution and error***	...the resolution is the smallest division which can be read easily. The resolution can be expressed as a percentage. *e.g. If the instrument can measure to 1 division and the reading is 10 divisions, the error can be expressed as 10±1 or as a percentage error of 10%.*

7	***Specificity****	...an instrument must measure only what it purports to measure. *e.g.false positives on drug tests due to detection of a similar naturally occuring substance.*
8	***Instrument use***	...there is a prescribed procedure for using an instrument which, if not followed, will lead to systematic and / or random errors. *e.g. When measuring the temperature of a liquid, if one takes the thermometer out of the liquid to read the thermometer, this will lead to systematically low or high readings, compared to reading the thermometer immersed in the liquid. More specifically, there is a prescribed depth of immersion for some thermometers that takes account of the expansion or contraction of the glass and the mercury (or alcohol) that are not in the liquid being measured.*
9	***Human error***	...even when an instrument is chosen and used appropriately, human error can occur. *e.g. Scales on measuring instruments can easily be misread.*

6 Reliability and validity of a single measurement

0	***Introduction***	Any measurement must be reliable and valid. *A measurement, once made, must be scrutinised to make sure that it is a valid measurement; it is measuring what was intended, and that it can be relied upon. Repeating readings and triangulation, by using more than one of the same type of instrument or by using another type of instrument, can increase reliability.*
1	***Reliability***	…a reliable measurement requires an average of a number of repeated readings; the number needed depends on the accuracy required in the particular circumstances *e.g. the height from which a ball is dropped could be checked if it was important that the drop height was accurate.*
2	***Reliability***	...instruments can be subject to inherent inaccuracy so that using different instruments can increase reliability. *e.g. Measurement of blood alcohol level can be assessed with a breathalyser and cross checked with a blood test. Also, temperature can be measured with mercury, alcohol, and digital thermometers to ensure reliability.*
3	***Reliability***	...human error in the use of an instrument can be overcome by independent, random checks. *e.g. Spot checks of measurement techniques by co-workers are sometimes built into routine procedures.*
4	***Validity***	...measures that rely on complex or multiple relationships must ensure that they are measuring what they purport to measure. *e.g. is the colour change a measure of bacterial activity or might something else have caused it?*

7 The choice of an instrument for measuring a datum

0	***Introduction***	Measurements are never entirely accurate for a variety of reasons. *Of prime importance is choosing the instrument to give the accuracy and precision required; a proactive choice rather than a reactive discovery that it wasn't the right instrument for the job!*
1	***Trueness or accuracy****	...trueness is a measure of the extent to which repeated readings of the same quantity give a mean that is the same as the 'true' mean. *e.g. If the mean of a series of readings of the height of an individual pupil is 173 cm and her 'true' height, as measured by a clinic's instrument is 173 cm, the measuring instrument is 'true' .*

2	***Non-repeatability***	...repeated readings of the same quantity with the same instrument never give exactly the same answer. *e.g. Weighing yourself on a bathroom scale in different places on the bathroom floor, or standing in a slightly different position on the scales, will result in slightly differing measurements. It is never possible to repeat the measurement in exactly the same way.*
3	***Precision***	…(sometimes called "imprecision" in industry) refers to the observed variations in repeated measurements from the same instrument. In other words, precision is an indication of the spread of the repeated measurements around the mean. A precise measurement is one in which the readings cluster closely together. The less the instrument's precision, the greater is its uncertainty. A precise measurement may not necessarily be an accurate or true measurement (and vice versa). The concept of precision is also called "reliability" in some fields. A more formal descriptor or assessment of precision might be the range of the observed readings, the standard deviation of those readings, or the standard error of the instrument itself. *e.g. For bathroom scales, a precise set of measurements might be: 175, 176, 175, 176, and 174 pounds.*
4	***Reproducibility***	…whereas repeatability (precision) relates to the ability of the method to give the same result for repeated tests of the same sample on the same equipment (in the same laboratory), reproducibility relates to the ability of the method to give the same result for repeated tests of the same sample on equipment in different laboratories. *e.g. 'Round Robins' are often used to check between different laboratories. A standardised sample is sent to each lab and they report their measurment(s) and degree of uncertainty. Labs are then compared.*
5	***Outliers in relationships***	…outliers, aberrant or anomalous values in data sets should be examined to discover possible causes. If an aberrant measurement or datum can be explained by poor measurement procedures (whatever the source of error), then it can be deleted. *e.g. an anomalous bounce would be deleted if the cause of the anomaly was known, but if it could not be explained, then further bounces would be needed to see if it was part of the inherent variation.*

8 Sampling a datum

0	***Introduction***	A series of measurement of the same datum can be used to determine the reliability of the measurement *We use the term 'sampling' to mean any sub-set of a population. The population might be a species of animal or plant, or even the possible sites where gold might be found. We shall also take the population to mean the infinite number of repeated readings.*
1	***Sampling***	...one or more measurements comprise a sample of all the possible measurements that could be made. *e.g. The measurement of a single blade of grass is a sample of all the blades of grass in a field. Also, a single measurement of the bounce height of a ball is a sample of the infinite number of such bounces that could be measured.*
2	***Size of sample***	.. the number of measurements taken. The greater the number of readings taken, the more likely they are to be representative of the population. *e.g.repeated readings on a ammeter in a particular circuit are a sample of all possible readings. The more readings taken, the more the sample represents the population of all possible readings.*

3	***Reducing bias in sample / representative sampling***	…measurements must be taken using an appropriate sampling strategy, such as random sampling, stratified or systematic sampling so that the sample is as representative as possible. *To find the height of college students, tables of random numbers can be used to select students.*
4	***An anomalous datum***	…an unexpected datum could be indicative of inherent variation in the data or the consequence of a recognised uncontrolled variable *e.g. Continuing the above examples, a very small height may have been recorded from a child visiting the college and should not be part of the population being sampled; whereas a very low rebound height from a squash ball may occur as a result of differences in the material of the ball and is therefore part of the sample.*

9 Statistical treatment of measurements of a single datum

0	***Introduction***	A group of measurements of the same datum can be described in various mathematical ways. *The statistical treatment of a datum is concerned with the probability that a measurement is within certain limits of the true measurement. The following are some basic statistics associated with a single datum.*
1	***Range***	…the range is a simple description of the distribution and defines the maximum and minimum values measured. *e.g. Measuring the height of carbon dioxide bubbles on successive trials in a yeast experiment, the following measurements were recorded and ordered sequentially: 1.8, 1.9, 2.1, 2.1, 2.1, 2.3, 2.4, 2.4, 2.5, 2.5 and 2.6 cm. The range is 0.8 cm (2.6 - 1.8)*
4	***Mean***	...the mean (average) is the sum of all the measurements divided by the number of measurements. *e.g. Continuing the example above, the mean is 2.2 cm*

10 Reliability and validity of a datum

0	***Introduction***	A datum must have a known (or estimated) reliability and validity before it can be used in evidence. *Any datum must be subject to careful scrutiny to ascertain the extent to which it:* *- is valid: that is, has the value of the appropriate variable been measured? Has the parameter been sampled so that the datum represents the population?* *- is reliable: for example, des the datum have sufficient precision? The wider the confidence limits (the greater the uncertainty), the less reliable the datum.* *Only then can the datum be weighed as evidence. Evaluation a datum also includes evaluating the validity of the ideas associated with the making of a single measurement.*
1	***Reliability***	...a datum can only be weighed as evidence once the uncertainty associated with the instrument and the measurement procedures have been ascertained. *e.g. the reliability of the volume of water absorbed by a papertowel should be judged in terms of the uncertainty of the instruments used in the method and any slops and spillages that might have occurred.*
2	***Validity***	...that a measurement must be of, or allow a calculation of, the appropriate datum. *e.g. the time it takes for a shoe to be pulled 50 cm across a surface is an invalid measure of the force required to move the shoe.*

11 Design of investigations: Variable structure

0	***Introduction***	The design of an investigation requires variables to be identified (as Independent, dependent and controlled) and measured. *An investigation is an attempt to determine the relationship, or lack of one, between the independent and dependent variables, or between two or more sets of data. Investigations take many forms but all have the same underlying structure. By identifying and understanding the basic structure of an investigation in terms of variables and types of variables, we can begin to evaluate the validity of data.*
1	***The independent variable***	...the independent variable is the variable for which values are changed or selected by the investigator. *e.g. length of resistance wire is the independent variable changed by the investigator to see what affect it has. The aspect of a slope is the independent variable selected in the field to investigate whether aspect affects vegetation.*
2	***The dependent variable***	...the dependent variable is the variable the value of which is measured for each and every change in the independent variable. *e.g. Continuing the examples above, the resistance of the wire or the density of bluebells.*
3	***Correlated variables***	…in some circumstances we are looking for a correlation only rather than any implied causation *e.g. the number of birds in the garden and the number of cats in the area.*
4	***Categoric variables***	...a categoric variable has values which are described by labels. Categoric variables are also known as nominal data. *e.g. "type of sugar" has data values of "icing", "castor", "granulated", etc.*
5	***Ordered variables***	...an ordered variable has values which are also descriptions, labels or categories but these categories can be ordered or ranked. Measurement of ordered variables results in ordinal data. *e.g. The variable of "age of bird" has ordered values "fledgling", "immature" and "adult". The variable "density of barnacles" can be assigned numbers according to a scale, but the numbers are not intervals but an ordered variable.*
6	***Continuous variables***	...a continuous variable is one which can have any numerical value and its measurement results in interval data. *e.g. height, current, time.*
7	***Discrete variables***	...a discrete variable is a special case in which the values of the variable are restricted to integer multiples. *e.g. The number of layers used to insulate a cup.*

12 Design: Validity, 'fair tests' and controls

0 ***Introduction***

Uncontrolled variation can be reduced through a variety of techniques.

Fair tests and controls aim to isolate the effect of the independent variable on the dependent variable. Laboratory-based investigations, at one end of the spectrum, involve the investigator changing the independent variable and keeping all the controlled variables constant. This is often called 'the fair test', but it is no more than one of several valid designs. At the other end of the spectrum are field studies where many naturally changing variables are measured and correlations sought. For example, an ecologist might measure many variables in a habitat over a period of time. Having collected the data, correlations might be sought between variables such as day length and emergence of a butterfly, using statistical treatments to ensure validity. The possible effect of other variables can be reduced by only considering data where the values of other variables are the same or similar. In between these extremes, are many types of valid designs that involve different degrees of manipulation and control. Fundamentally, all these investigations have a similar structure; what differs are the strategies to ensure validity.

1 ***Fair test***

...a fair test is one in which only the independent variable has been allowed to affect the dependent variable.

e.g. A laboratory experiment about the effect of temperature on dissolving time, where only the temperature is changed. Everything else is kept exactly the same.

2 ***Control variables in the laboratory***

...other variables can affect the results of an investigation unless their effects are controlled by keeping them constant.

e.g. In the above experiment, the mass of the chemical, the volume of liquid, the stirring technique, and the room temperature are some of the variables that should be controlled.

3 ***Control variables in field studies***

...some variables cannot be kept constant and all that can be done is to make sure that they change in the same way.

e.g. In a field study on the effect of different fertilisers on germination, the weather conditions are not held constant but each experimental plot is subjected to the same weather conditions.

4 ***Control variables in surveys***

…the potential effect on validity of uncontrolled variables can be reduced by selecting data from conditions that are similar with respect to other variables.

e.g. In a field study to determine whether light intensity affects the colour of 'dog's mercury leaves (Mercurialis perrenis, a woodland plant), other variables are recorded, such as soil nutrients, pH, and water content. Correlations are then sought by selecting plants growing where the value of these variables is similar. Also, in a survey investigating if blood pressure increased with age, other variables such as diet, fitness, weight to height ratio, etc. are recorded.

5 ***Control group experiments***

…control groups are used to ensure that any effects observed are due to the independent variable(s) and not some other unidentified variable. They are no more than the default value of the independent variable.

e.g. seedlings are grown in normal soil (control group) as well as in a variety of composts with different nutrients to ensure that the batch of seeds was viable and the temperature and watering regime etc were suitable for normal growth.

13 Design: Choosing values

0	***Introduction***	Choosing the values of the variables in an investigation. *The values of the variables need to be chosen carefully. This is possible in the majority of investigations, prior to the data being collected. In field studies where data are collected from variables that change naturally, some of these concepts can only be applied retrospectively.*
1	***Trial run***	…a trial run can be used to establish the broad parameters required of the experiment (scale, range, number) and help in choosing instrumentation and other equipment *e.g. paper helicopters of different dimensions are dropped from different heights to determine which is likely to be suitable for the investigation.*
2	***The sample***	…issues of sample size and representativeness apply in the same way as in sampling a datum (Measurement section). *e.g. This cannot be determined in advance, or be a set number, if the inherent variation is unknown.*
3	***Relative scale***	...the choice of sensible values for quantities is necessary if measurements of the dependent variable are to be meaningful. *e.g. dropping a large paper helicopter from 0.5m doesn't allow it to start to turn which will invalidate the results.*
4	***Range***	...the range over which the values of the independent variable is chosen is important in ensuring that any pattern is detected. *e.g. An investigation into the effect of competition for light with wheat seeds placed between 5 and 10cm apart would show little effect of crowding.*
5	***Interval***	...the choice of interval between values determines whether or not the pattern in the data can be identified. *e.g. An investigation into the effect of temperature on enzyme activity would not show the complete pattern if 20°C intervals were chosen.*
6	***Number***	...a sufficient number of readings is necessary to determine the pattern. *e.g. The number is determined partly by the range and interval issues discussed above, but in some cases for the complete pattern to be seen, more readings may be necessary in one part of the range than another. This applies particularly if the pattern changes near extreme values, for example, in a spring extension experiment at the top of the range of the mass suspended on the spring.*

14 Design: Accuracy and precision

0	***Introduction***	Ensuring appropriate accuracy and precision. *The design of the investigation must provide data with sufficiently appropriate accuracy and precision to answer the research question. This consideration should be built into the design of the investigation. Different investigations will require different levels of accuracy and precision depending on their purpose.*
1	***Determining differences***	...there is a level of precision which is sufficient to provide data which will allow discrimination between two or more means. *e.g. Which is the most absorbant 'economy paper towel'? When the differences are potentially small the degree of precision in the measurement method must be sufficient to enable differences to be attributed to the paper towel rather than uncertainty in the measurements.*

2 ***Determining patterns*** ...there is a level of precision which is required for the trend in a pattern to be determined.

e.g. Large error-of-measurement bars on the points of a line graph may not allow discrimination between an upward curve or a straight line.

15 Design: Tables

1 ***Tables*** ...tables can be used as organisers for the design of an experiment by preparing the table in advance of the whole experiment. A table has a conventional format.

e.g. An experiment on the effect of temperature on the dissolving time of calcium chloride:

16 Reliability and validity of the design

0 ***Introduction*** An evaluation of an investigation must consider reliability and validity.

In evaluating the design of an investigation, there are two overarching questions:

- will the measurements result in sufficiently reliable data to answer the question?
- will the design result in sufficiently valid data to answer the question?

Evaluation the design of an investigation included evaluating the reliability and validity of the ideas associated with the making of a single measurement and with each and every datum.

1 ***Reliability of the design*** ...the reliability of the design includes a consideration of all the ideas associated with the measurement of each and every datum.

e.g. Factors associated with the choice of the measuring instruments to be used must be considered, for instance, the error associated with each measuring instrument. The sampling of each datum and the accuracy and precision of the measurements should also be considered. This includes the sample size, the sampling technique, relative scale, the range and interval of the measurements, the number of readings, and the appropriate accuracy and precision of the measurements.

2 ***Validity of the design*** ...the validity of the design includes a consideration of the reliability (as above) and the validity of each and every datum.

e.g. This includes the choice of measuring instrument in relation to whether the instrument is actually measuring what it is supposed to measure. This includes considering the ideas associated with the variable structure and the concepts associated with the fair test. For instance, measuring the distance travelled by a car at different angles of a ramp will not answer a question about speed as a function of angle.

17 Data presentation

0 Data can be presented in a number of ways.

Having established that the design of an investigation is reliable and valid, what do we need to understand to explore the relationship between one variable and another? Another way of thinking about this is to think of the pattern between two variables or two sets of data. What do we need to understand to know that the pattern is valid and reliable? The way that data are presented allows patterns to be seen. There is a close link between graphical representations and the type of variable they represent.

1	***Tables***	…a table is a means of reporting and displaying data. But a table alone presents limited information about the design of an investigation e.g. control variables or measurement techniques are not always overtly described. *e.g. a table showing type of sugar and dissolving time does not indicate the mass of sugar used, the volume or temperature of the water and whether it was stirred and how vigorously.*
2	***Bar charts***	...bar charts can be used to display data in which the independent variable is categoric and the dependent variables is continuous. *e.g. The number of pupils who can and cannot roll their tongues would be best presented on a bar chart.*
3	***Line graphs***	...line graphs can be used to display data in which both the independent variable and the dependent variable are continuous. They allow interpolation and extrapolation. *e.g. The length of a spring versus the force applied would be best displayed in a line graph.*
4	***Scatter graphs (or scatter plots)***	... are used to display data in which both the independent variable and the dependent variable are continuous. Scatter graphs are often used where there is much fluctuation in the data because they can allow an association to be detected. Widely scattered points can show a weak correlation, points clustered around, for example, a line can indicate a relationship. *e.g. The dry mass of the aerial parts of a plant and the dry mass of the roots.*
5	***Histograms***	...histograms can be used to display data in which a continuous independent variable has been grouped into ranges and in which the dependent variable is continuous. *e.g. On a seashore, the distance from the sea could be grouped into ranges, i.e. lower, middle and upper shore, and the number of limpets in each range plotted in a histogram.*
8	***Other forms of display***	...data can be transformed, for example, to logarithmic scales so that they meet the criteria for normality which allows the use of parametric statistics. *e.g. coloured shading or symbols could be used on a map to indicate the density of plants in an area of woodland. A pictorial scaled drawing could be used to show the amount of emissions from different forms of transport.*

19 Patterns and relationships in data

0	***Introduction***	Data must be inspected for underlying patterns. *Patterns cannot be treated in isolation from the physical system that they represent, because patterns represent the behaviour of variables in that system. Patterns can be seen in tables or graphs or can be reported by using the results of appropriate statistical analysis. The interpretation of patterns and relationships must respect the limitations of the data. For instance, there is a danger of over-generalizing or of implying causality when there may be a different, less direct type of association.*

1	***Types of patterns***	…there are different types of association such as causal, consequential, indirect or chance associations. "Chance association" means that observed differences in data sets, or changes in data over time, happen simply by chance alone. We must sceptically be open to possibility that a pattern has emerged by chance alone. Statistical tests give us a rational way to estimate this chance. *e.g. In any large multivariate set of data, there will be associations, some of which will be chance associations. Even if x and y are highly correlated, x does not necessarily cause y: y may cause x or z may cause x and y. (See 1.3.) Also, changes in students' understanding before and after an intervention may not be significant and/or may be due to other factors.*
2	***Linear relationships***	...straight line relationships (positive slopes, negative, and vertical and horizontal as special cases) can be present in data in tables and line graphs and that such relationships have important predictive power (y = mx + c) *e.g. Height and time for a falling object.*
3	***Proportional relationships***	… direct proportionality is a particular case of a straight line relationships with consequent predictive characteristics. The relationship is often expressed in the form (y = mx). *e.g. Hooke's law: the length of a spring is directly proportional to the force on the spring.*
4	***'Predictable' curves***	…patterns can follow predictable curves (y=x2 for instance), and that such patterns are likely to represent significant regularities in the behaviour of the system (velocity against time for a falling object for instance) *e.g. Velocity against time for a falling object. Also, the terminal velocity of a parachute against its surface area.*
5	***Complex curves***	.. some patterns can be modelled mathematically to give approximations to different parts of the curve (Hooke's law for a spring taken beyond its elastic limit for instance) *e.g. Hooke's law for a spring taken beyond its elastic limit.*
6	***Empirical relationships***	… patterns can be purely empirical and not be easily represented by any simple mathematical relationship. *e.g. number of greenfly on a rose bush over time.*
7	***Anomalous data***	… patterns in tables or graphs can show up anomalous data points which require further consideration before excluding them from further consideration. *e.g. A 'bad' measurement or datum due to human error.*
8	***Line of best fit***	…for line graphs (and scatter graphs in some cases) a 'line of best fit' can be used to illustrate the underlying relationship, 'smoothing out' some of the inherent (uncontrolled) variation and human error

20 Reliability and validity of the data in the whole investigation

0	***Introduction***	An overall solution to a problem can included repeated experiments and triangulation from other data sources. *So far we have considered the data within a single investigation. In reality the results of an investigation will usually be compared with evidence from other investigations.* *In evaluating the whole investigation, all the foregoing ideas about evidence need to be considered in relation to the two overarching questions:* *- are the data reliable?* *- are the data valid?* *In addressing these two questions, ideas associated with the making of single measurements and with each and every datum in an investigation should be considered. The evaluation should also include a consideration of the design of an investigation, as well as ideas associated with measurement, with the presentation of data, and with the interpretation of patterns and relationships.*
1	***A series of experiments***	...a series of experiments can add to the reliability and validity of evidence even if, individually, their precision does not allow much weight to be placed on the results of any one experiment alone.
2	***Secondary Data***	...data collected by others is a valuable source of additional evidence, provided its value as evidence can be judged. *e.g. comparing the results with data reported from other sources*
3	***Triangulation***	...triangulation with other methods can strengthen the validity of the evidence.

21 Relevant societal aspects

0	***Introduction***	Evidence must be considered in the light of personal and social experience and the status of the investigators. *If we are faced with evidence and we want to arrive at a judgement or decision that leads to action, other factors outside the domain of science may become relevant, some of which are listed here.*
1	***Credibility of evidence***	... credibility has a lot to do with face validity: consistency of the evidence with conventional ideas, with common sense, and with personal experience. Credibility increases with the degree of scientific consensus on the evidence or on theories that support the evidence. Credibility can also turn on the type of evidence presented, for instance, statistical versus anecdotal evidence. *e.g. Evidence showing low emissions of dioxins from a smokestack is compromised by photos of black smoke spewing from the smokestack (even though dioxins are relatively colourless). Also, concern for potential health hazards for workers in some industries often begins with anecdotal evidence, but is initially rejected as not being scientifically credible.*
2	***Practicality of consequences***	... the implications of the evidence may be practical and cost effective, or they may not be. The more impractical or costly the implications, the greater the demand for higher standards of validity and reliability of the evidence. *e.g. The negative side effects of a drug may outweigh its benefits, for all but terminally ill patients. Also, when judging the evidence on the source of acid rain, Americans will likely demand a greater degree of certainty of the evidence than Canadians who live down wind, because of the cost to American industries to reduce sulphur dioxide emissions.*

3	***Experimenter bias***	...evidence must be scrutinized for inherent bias of the experimenters. Possible bias may be due to funding sources, intellectual rigidity, or an allegiance to an ideology such as scientism, religious fundamentalism, socialism, or capitalism, to name but a few. Bias is also directly related to interest: Who benefits? Who is burdened? *e.g. Studying the link between cancer and smoking funded by the tobacco industry; or studying the health effects of genetically modified foods funded by Green Peace. Also, the acid rain issue (above) illustrates different interests on each side of the Canadian/American border.*
4	***Power structures***	...evidence can be accorded undue weight, or dismissed too lightly, simply by virtue of its political significance or due to influential bodies. Trust can often be a factor here. Sometimes people are influenced by past occurrences of broken trust by government agencies, by industry spokespersons, or by special interest groups. *e.g. Studies published in the New England Journal of Medicine tend to receive greater weight than other studies. Also, the pharmaceutical industry's negative reaction to Dr. Olivieri's research results that were not supportive of their drug Apotex at Toronto's Hospital for Sick Children in 2001*
5	***Paradigms of practice***	...different investigators may work within different paradigms of research. For instance, engineers operate from a different perspective than scientists. Thus, evidence garnered within one paradigm may take on quite a different status when viewed from another paradigm of practice. *e.g. Theoretical scientists tend to use evidence to support arguments for advancing a theory or model, whereas scientists working for an NGO, for instance, tend to use evidence to solve a problem at hand within a short time period. Theoretical scientists have the luxury of subscribing to higher standards of validity and reliability for their evidence.*
6	***Acceptability of consequences***	…Evidence can be denied or dismissed for what may appear to be illogical reasons such as public and political fear of its consequences. Prejudice and preconceptions play a part here. *.e.g. During the tainted blood controversies in the mid 1980s, the Canadian Red Cross had difficulty accepting evidence concerning the transmission of HIV in blood transfusions. BSE and traffic pollution are examples in Europe.*
7	***Status of experimenters***	...the academic or professional status, experience and authority of the experimenters may influence the weight which is placed on the evidence. *e.g. Nobel laureates may have their evidence accepted more easily than new researchers' evidence. Also, A botanist's established reputation affects the credibility of his or her testimony concerning legal evidence in a courtroom.*
8	***Validity of conclusions***	…conclusions must be limited to the data available and not go beyond them through inappropriate generalisation, interpolation or extrapolation *e.g. The beneficial effects of a pharmaceutical may be limited to the population sample used in the human trials of the new drug. Also, evidence acquired from a male population concerning a particular cardiac problem may not apply as widely to a female population.*