QUALITY IN STUDY SUPPORT QiSS

DRAWING CONCLUSIONS FROM DATA

SUMMARY
This guide seeks to aid schools in drawing valid conclusions from the data collected. It looks at identifying trends, assessing reliability and validity of data and conclusions, and whether or not statistical analysis is necessary.

INTRODUCTION

Once you have collected your data, you need to look at what it really shows you. Sometimes (but very rarely) the conclusion is obvious. If every single member of a reading club made a huge gain in reading age and no member of the control group made any such gain, then you can be reasonably confident that the reading club benefited the pupils' reading ability! In the real world, however, data is rarely that clear cut. There will usually be variation between individuals and the degree of this may shed doubt on your conclusions. If an improvement is shown, you need to judge if this improvement is significant.

Even if a significant effect is noticed, you need to consider the reliability of your conclusion - how certain can you be that what you did caused the effect, rather than any other factors?

This guide takes you through the necessary steps to ensure that you get real value and meaning from your collected results. It deals with quantitative data, as it is often difficult to draw general conclusions from qualitative data.

IDENTIFYING TRENDS

You hope that attendance at study support in general, or at a particular activity, will have some measurable effect on the participants. Trends are important when you want to study the effect of *different amounts* of study support.
Examples might be:

- The effect of different levels of attendance at study support on SAT level achieved at the end of a Key Stage.
- The effect of different number of attendances at a Maths club on Maths results at the end of a year.
- Progress of a measurable sports skills during the course of an extended programme.

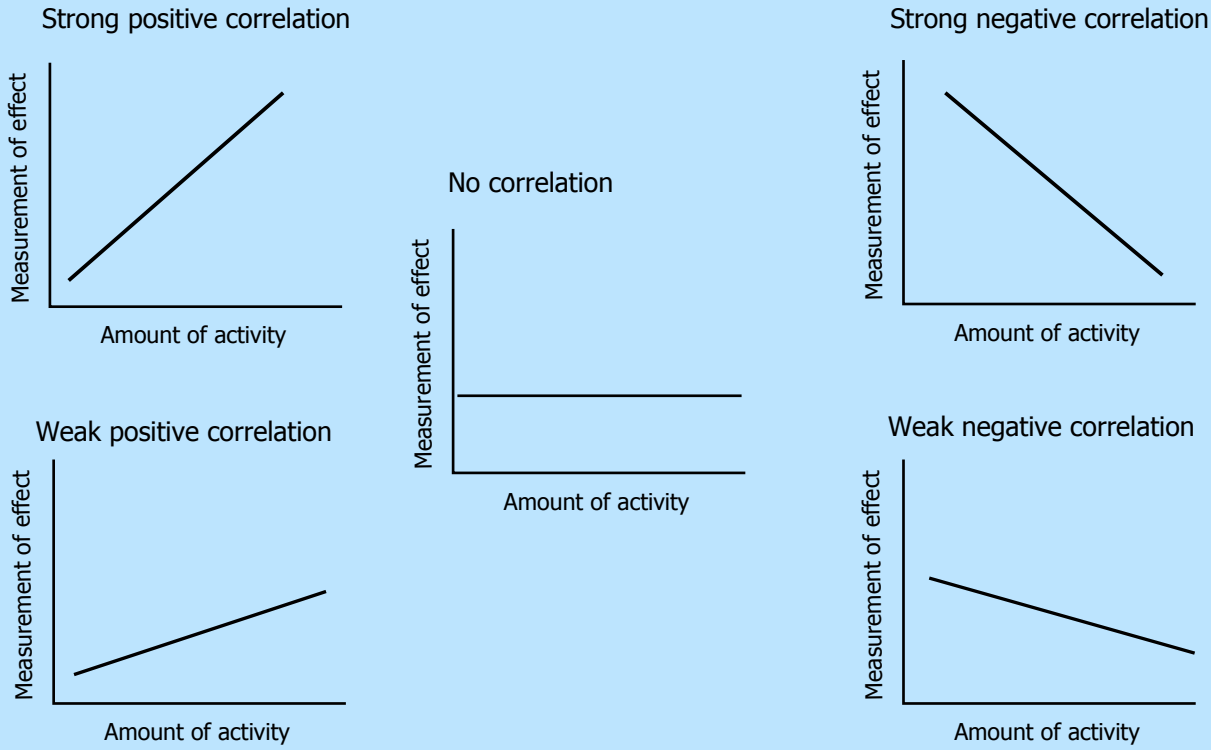
What you are looking for is some sort of relationship between what you provided and a measurable outcome. This is called a correlation, and is best shown by plotting the effects on a graph- different types of correlation are shown in Box 1.

Strong positive correlation
Increasing the amount of activity increases the measurable outcome considerably.
Weak positive correlation
Increasing the amount of activity increases the measurable outcome to some extent.
No correlation
The activity has no measurable effect on the outcome
Strong negative correlation
The activity significantly decreases the measured outcome (this can be desirable - e.g. If the activity was a 'health club' and the outcome was fast food items eaten in a week).

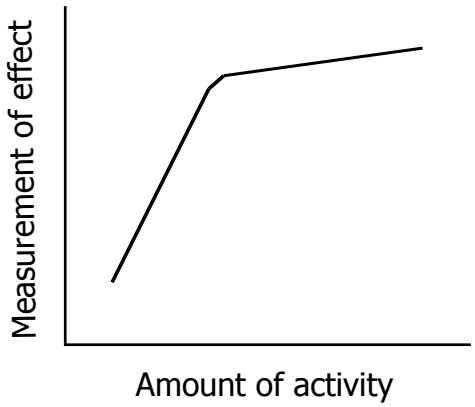
Weak negative correlation

Increasing the amount of activity decreases the measurable outcome to some extent.

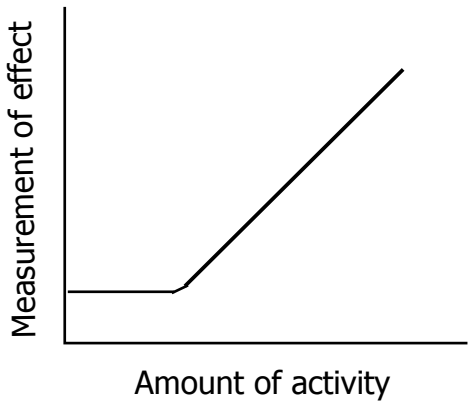
Box 1 - Types of correlation



Very often, the points on the graph will not fall in an absolutely regular pattern. The lines drawn on the graphs in Box 1 simply indicate general trends in the data. Other patterns are possible:



This type of trend indicates that the activity has an initial effect, but that after a while the rate of improvement becomes much less. This could be useful in determining the optimum length for the activity.



This pattern of trend indicates that improvement takes a while to 'kick in'. This also gives useful information, this time on the *minimum* duration of the activity.

It can be seen that analysing such trends can provide a range of useful information, beyond the simple question of "Does this activity have an effect?"

ACCURACY AND RELIABILITY OF DATA

It is unwise to draw firm conclusions from inaccurate or unreliable data. Accurate data reflects reality - it is near to the 'true' value of what you are measuring. Often, there is little problem with the accuracy of data collected in connection with study support. Unless there is human error, for instance, it is impossible to be 'inaccurate' when recording attendance data - either a pupil is there or is not!
Accuracy can sometimes become an issue with questionnaires and evaluation forms, however. Some common problems are:

1) The pupils do not understand the question
2) The pupils misunderstand the question - this can often occur when asked to give a rating of 1-5 for a response, for instance. Some pupils may 'reverse' the scale (e.g. thinks 1 means
'good', when in fact it means 'bad')
3) The pupils tend to give the response that they feel is 'expected' rather than a genuine opinion.
4) The pupils answer in 'friendship groups', all of whom give the same response.
5) Self reporting of progress *can* be subjective and inaccurate. This is particularly the case with pupils of low self-esteem, who can either play down their progress or be reluctant to admit to anything less than absolute success.

Careful planning of the questionnaire and its administration, and careful observation of the responses can help to avoid such inaccuracies. If a particular inaccuracy is likely to skew the results, some data may need to be excluded. In the suggested scenario above, for instance, where a pupil may be following his or her friend, there *may* be a case for recording their two responses as a single response. This exclusion of results should only be done where there is a clear indication of inaccuracy. The researcher must not exclude data simply on the basis that it is unexpected or does not fit in with a particular hypothesis.

Box 2: Illustration of reliable and unreliable results

The results below show a comparison of the improvement of actual GCSE grades (indicated as +/- number) over predicted grades, in pupils who did or did not regularly attend study support

	Grade improvement or decrease over predicted grade										Avge
Attended	+3	0	+3	-1	-1	-1	+2	+2	-2	+2	+0.8
Did not attend	0	+1	+2	0	-1	+2	-1	0	+1	+2	+0.5

The results above show an apparent slight increase in GCSE performance by the group who attended, but the results are very variable and several pupils who attended did worse than some who did not. These results are rather unreliable

	Grade improvement or decrease over predicted grade										Avge
Attended	+2	0	+1	+1	+1	+1	0	0	+1	+1	+0.8
Did not attend	+1	0	+1	0	-1	-1	+2	+1	+2	0	+0.5

The results here produce the same averages but the figures for the attendees are much less variable and can therefore be considered more reliable. The figures for those who did not attend show a little more variability than for those who did.

Note that both sets of data in Box 2 show a relatively small improvement in those who did attend study support compared to those who did not. A judgement will still have to be made as to whether such an improvement can be judged *significant* (see below), but that judgement can be made much more easily with reliable results than with unreliable ones. If the difference between groups is very large, reliability of results, while still desirable, becomes less important.

In order to maximise reliability, the following measures are suggested:

1) Ensure that results are as accurate as possible (see above). Inaccurate methods can increase variation.
2) Ensure sample size is big enough ('rogue' results can have a big effect in small samples).

If your results are not as reliable as you would like, it does not mean that they are useless and should be discarded. Unreliability can make any conclusions a little more tentative, but not necessarily invalid.

RELIABILITY OF CONCLUSIONS

Research data rarely proves or disproves a given idea. It tends to provide evidence, rather than proof. People inexperienced in data handling sometimes draw conclusions which go beyond what their data actually show.

The reliability of any conclusions drawn will increase if the actual data is reliable and accurate (see above). Conclusions will be strengthened when trends are very pronounced. Some small differences may not be significant, although they may be if they show real consistency. For example, let us suppose that a group of pupils who have attended a Science Club score, on average, one more mark on a science test than those who did not. This difference is probably not significant. If, however, it is found that the one mark increase is repeated in *all* of the twenty science tests they have during the course of the year, then it is possible that the benefit *is* significant, even if small.

A simple way to judge the significance of any difference is to analyse the results statistically (see below).

It is difficult to draw reliable conclusions unless some form of control group is used. A 'Reading Buddy' scheme may appear to produce a significant rise in the participants' reading age. However, unless you know that a group of similar pupils who did not attend the scheme showed much less advancement, the conclusion is unreliable. Setting up a control group is sometimes difficult, and this is covered in another guidance leaflet in this series, but it is usually essential in order to draw reliable and valid conclusions. Sometimes, national data may be a substitute for a school based control group.

Ensure that any conclusions drawn do not go beyond what the data actually indicates. In particular, data will very often indicate an effect but not provide any evidence for the cause of that effect. For example, let us suppose that attendance at a Summer transition course resulted in the pupils who attended reporting less anxiety in the first week of their new school year than those pupils who did not attend. It is reasonable to conclude that attendance was linked with less anxiety, but the data is unlikely to give precise evidence as to why this occurred. A conclusion like "having the chance to meet their peers and teachers in the transition course caused them to feel less anxious when they started school" is not justified (unless that was specifically identified as a factor by the pupils concerned). It may be, for instance, that those pupils who were willing to attend such a course were inherently less anxious to begin with.

When drawing conclusions, it is always worthwhile to think of as many reasons as possible for the effect noticed, not just the one you may consider 'obvious'.

WHEN TO USE STATISTICS

Many impact measurements are not 'research studies' and it is understandable if teachers do not wish to get into statistical analysis. Statistical analysis is advisable when:

- The significance of the results is not certain
- The outcome of the study has important implications

A later guide in this series deals with the decision on whether to use statistics or not in much more detail.

SUMMARY - QUESTIONS TO ASK

- Is my data accurate?
- Is my data reliable?
- Is there an adequate control?
- Does my data show a clear trend?
- Are any differences significant?
- Do I need to do statistical tests in order to find out?
- Is my conclusion limited to what the data actually shows?
- Are other interpretations possible?