



# SAMPLE SIZES AND CONTROL GROUPS

## SUMMARY

Any conclusions drawn from measuring impact are more reliable if the sample size is adequate and there is a control group. On occasions, one or both of these factors may not be important. This pamphlet gives guidance on whether and how to judge sample size, and control group quality

## INTRODUCTION

Impact can be measured in various ways and in various contexts. Sometimes, it can involve a good deal of effort to collect and analyse data. It is important, therefore, that this time spent results in conclusions that are valid and reliable. Two ways to ensure this are by having a big enough sample, and a 'control' group with which to compare. These factors are not always necessary though, depending on the type of measurement being made

### WHEN IS SAMPLE SIZE IMPORTANT?

Sample size becomes a factor in two situations:

- When you are trying to establish some general trend
- When you are making a comparison between two different sets of data

Examples of these types of situations are:

- Does attendance at the study support programme improve attainment?
- Is our new study support programme better than the old one?
- Does attendance at our cooking club lead to participants having a healthier diet?
- Does attendance at homework club improve classroom performance?
- Does our leadership course improve self-confidence?
- Is there any difference in the effects of our programme between boys and girls?

Sample size is not a factor when you are focusing on the specific outcomes of an activity on individuals or a small group, such as:

- Do the pupils in our karate club progress to the next belt during the course?
- Do the pupils in the pottery class all produce a pot without help by the end of the course?
- Has pupil X's attitude to school improved after attending a residential week?

### WHY IS SAMPLE SIZE IMPORTANT?

The behaviour of human beings is very complex, and individual reactions to any set of circumstances will vary. Whatever is measured, therefore, there will be variation between individuals. If there is only a small sample, then people with 'unusual' reactions can skew the results, but in larger samples this will not happen.

As an example, School A was monitoring the effect of a revision skills course on GCSE grades. 14 pupils attended the course, but, when impact was assessed, the improvement of GCSE grades was

marginal and not apparently significant. However, two of the pupils had attended the sessions because their friends were there, and had not shown any interest in taking part in the activities. Whilst all the other pupils had made significant improvement with their GCSE grades, these two actually performed worse than expected. Because they represented a fairly high proportion of the group (14%) and because they did significantly worse than expected, the results were skewed. If these two pupils had been in a group of 40, their effect would have been much less and probably insignificant. (There would be a case for ignoring these two pupils, of course, if the person analysing the data knew of the situation).

So, sample size needs to be big enough to overcome the effect of occasional ‘rogue’ results.

### WHAT IS A BIG ENOUGH SAMPLE?

How long is a piece of string? The answer to the sample size question is really “as many as possible”, but there are often limits to what can be achieved, and in many cases the numbers are not under your control (you can’t force more people to attend an activity so that the sample size is sufficient!).

*Research Methods in Education* (Cohen et al., 2007) suggests that if general conclusions about the school population as a whole are to be reached, the sample size needs to be at least 30.

### WHAT IF I CAN'T GET A BIG ENOUGH SAMPLE?

In some situations it will be impossible to get a group of 30 or more to analyse. This does not mean that measuring impact will be a complete waste of time, however. It simply means that you can be less sure of your conclusions. Let us re-visit the ‘Revisions Skills’ scenario, above. If 20 pupils attend, and the average improvement over predicted GCSE grades is 1 grade per subject, then the scale of the improvement and its consistency will overcome any doubts due to sample size (particularly if a control group showed no improvement).

Small sample size, therefore, is simply a ‘factor of doubt’ which needs to be set alongside other evidence.

### WHAT IS A CONTROL GROUP AND WHY DO I NEED ONE?

A control group is a group similar to the experimental group, but where the factor thought to be causing an effect is removed.

For example, suppose a school wants to look at the effect of a homework club on classroom performance. Regular attenders seem to have improved considerably in classroom tests between September and January. This does not mean, however, that attendance at homework club caused this. It may be that increasing maturity had an effect, or that the tests carried out in January were relatively easier than in September, or that some other factor (e.g. New and effective behaviour management system) had an effect.

In order to isolate the effect of the homework club, regular attenders need (ideally) to be compared with a group of non-attenders, and this latter group should be *as similar as possible* to the homework club group, for example:

- Similar gender balance
- Similar attendance at school
- Similar spread of family background
- Similar spread of ability
- Same teachers
- Same tests taken (in September and January)

It is very unlikely that all of these criteria can be met, but if they could and attenders outperformed non-attenders, then there could be no doubt that the homework club benefited performance. Without any control group, there are too many other possible factors to be confident in your conclusion.

### WHEN IS A CONTROL GROUP UNNECESSARY?

You don't need a control group when you are comparing two experimental groups, for example, if you wanted to see if a particular activity benefited boys and girls equally, then you do not need (and it would be impossible to have) a control group that was neither male nor female!

Sometimes. Common sense must be used. If you want to see if attendance at an archery club improved participants' accuracy in hitting the target, then it hardly needs checking whether pupils who do no archery get better at it naturally! It is theoretically possible that it *might* happen, but we have better ways to spend our time than to check on it in that scenario.

#### Examples of the use of control groups

A school wanted to see if a year spent doing the Duke of Edinburgh award course had any effect of physical fitness. Pupils did fitness assessments in PE lessons at various times of the year. The group doing the DofE award scored better in fitness assessment in June compared with the previous September. In order to validate the apparent effect, a control group was selected. The following factors were considered:

1. Same school year groups (and same number of pupils in each year group)
2. Same gender balance
3. Similar spread of fitness scores in the September test

Given the fact that every pupil did these fitness tests, these factors were relatively simple to control. Other factors such as activities outside school were considered too difficult to collect data on and were ignored. Although this meant the control group was not perfect (it never will be) it is quite adequate for the purpose.

---

A school wanted to see if attendance at study support improves attainment at Key Stage 3 and found that those that attended study support did improve their levels significantly. However, it may be that more motivated pupils attend study support and also improve because of their general level of motivation. In other words, cause and effect are not established. It would be better to look for a control group consisting of pupils who appear to be well motivated but do not attend study support, and compare them with pupils of similar ability in the study support group.

### WHAT IF I CAN'T SET UP A PROPER CONTROL GROUP??

No control group will be perfect but any type of control group will be better than having none (assuming the study needs one). The weaker the control group, the more difference will be needed to draw confident conclusions.

Sometimes, no control group can be established. In some primary schools, for instance, virtually every pupil attends study support and so a control group cannot be set up as a comparison with attenders. This seriously weakens, but does not invalidate conclusions.

## ALTERNATIVES TO CONTROL GROUPS

Sometimes, other data can be used instead of a control group. For instance, if studying progress in reading age over a year in a reading club, comparison can be made to figures on average progression to establish what is 'normal'. The purpose of a control group is to judge natural progression, and in some case national data may be available to use.

## SUMMARY

When measuring impact:

1. Decide if sample size is important
2. If it is, try to a sample of at least 30
3. If you can't, bear this in mind when drawing conclusions
4. Decide if a control group is necessary
5. If it is, try to make it as similar as possible to the experimental group
6. Depending on the existence or quality of a control group, adjust your opinion on the validity and reliability of the conclusion.