SSERC

Evidence Based Education

A guide for teachers considering carrying out their own school based research

Jim Stafford

Foreword

Amid all the changes and developments in Scottish education in recent years, one of the most significant has been the reconceptualisation of what it means to be a teacher in Scotland in the 21st century.

Lying at the heart of this has been the importance of teachers as "enquiring practitioners" who engage in and with educational research. Graham Donaldson's seminal report *Teaching Scotland's Future* points out that:

> *"The most successful education systems invest in developing their teachers as reflective, accomplished and enquiring professionals who are able, not simply to teach successfully in relation to current external expectations, but who have the capacity to engage fully with the complexities of education and to be key actors in shaping and leading educational change."*

The notion of teachers as "enquiring practitioners" is not new. Over many years, terms such as "action research", "teacher research" and "collaborative enquiry" have all been used to describe the kinds of professional learning activities that support teachers to become more engaged with research to enhance their own learning, that of colleagues and ultimately pupil experiences.

Teachers being and becoming enquiring practitioners is clearly evident in GTC Scotland's Professional Standards. The Standards for Provisional and Full Registration expect that teachers have knowledge and understanding of the importance of research and engagement in professional enquiry. More specifically, they expect registered teachers to know how to engage critically in enquiry, research and evaluation, individually or collaboratively, and apply this in order to improve learning and teaching. Within the Standard for Career-Long Professional Learning, "Enquiry and Research" is one of the key areas of professional learning, requiring teachers to exhibit the following professional actions:

- develop and apply expertise, knowledge and understanding of research and impact on education;
- develop and apply expertise, knowledge, understanding and skills to engage in practitioner enquiry to inform pedagogy, learning and subject knowledge;
- lead and participate in collaborative practitioner enquiry.

Professional learning is what teachers engage in to stimulate their thinking and professional knowledge and ensure that their practice is critically informed and up-to-date. Opportunities to engage in such learning have been supported by the introduction of Professional Update and by the easy access to high quality academic and practitioner research provided through GTC Scotland's EBSCO resources and its Research Hub, both available via MyGTCS on the GTC Scotland website.

Cover image (bottom right) by Janoon028/FreeDigitalPhotos.net

It is well recognised that by undertaking a wide range of high-quality, sustained professional learning experiences, teachers are more likely to inspire pupils and provide high quality teaching and learning experiences, enabling learners to achieve their best. This takes on an even greater importance as teachers in Scotland address Scotland's Attainment Challenge and the priorities set out in the National Improvement Framework, both of which focus on closing the attainment gap.

Engagement in conducting individual or collaborative research with colleagues provides a rich opportunity for teachers to develop and enhance their professional knowledge and practice in order to progress the quality of learning and teaching and school improvement.

Given this background, SSERC's guide for teachers considering carrying out their own school-based research written by Jim Stafford is very timely. In it, he sets out clearly different approaches to conducting educational research, with each chapter containing thought-provoking case studies and eminently practical advice. Jim's chapter on "Designing your own" aligns well with the increasing encouragement and support being given to teachers to engage in and with their own research at a practical level. It reminds us that all education research should have a purpose and provide evidence for teachers taking decisions that will lead to improvement in practice and/or outcomes for learners. It also emphasises that research conducted by teachers doesn't need to be ground breaking or Earth shattering and that, properly conducted, it can be a powerful tool in delivering change in practice and improvement.

Ken Muir
Chief Executive
The General Teaching Council for Scotland

June 2016

Preface

Decision making in education should be based on sound research evidence and should be the concern of every teacher.

Learning how research works is important so that teachers become research literate enabling them both to carry out their own research into what is effective and to be critical consumers of the research findings of others. Educational initiatives that are not supported by evidence should be challenged and the research that is required to demonstrate their effectiveness should be identified.

"I think there is a huge prize waiting to be claimed by teachers. By collecting better evidence about what works best, and establishing a culture where this evidence is used as a matter of routine, we can improve outcomes for children, and increase professional independence."

Ben Goldacre - doctor, epidemiologist and campaigning journalist.

Contents

		Page
	Introduction	**1**
	Case study box 1: A badly conducted trial and how it could have been sorted	*2*
1	**Randomised controlled trials - the gold standard of evidence**	**4**
1.1	The placebo effect, blind and double blind trials	4
1.2	Historical and other similar types of control	5
	Case study box 2: Can we compare the results of National 5, Intermediate 2 and Standard Grade?	*6*
	Case study box 3: Base line studies	*7*
1.3	Randomisation	8
	1.3.1 Blocking	*8*
	1.3.2 Paired designs	*9*
	1.3.3 Within subject designs	*9*
2	**Correlation studies**	**10**
	Case study box 4: Can drinking water improve exam performance?	*11*
2.1	Case control studies	12
	Case study box 5: Are Easter school revision classes and study clubs any good?	*12*
3	**Systematic reviews and meta-analyses**	**13**
4	**Social research methods**	**15**
	Case study box 6: How school inspections work	*16*
4.1	Questionnaires	16
4.2	Interviews	17
4.3	Observation	18

Page

5	**Analysing results**	**19**
	Case study box 7: Using statistics to analyse results	*20*
	Case study box 8: Should primary teachers have a pass in Higher Mathematics?	*21*
	Case study box 9: Change for change's sake?	*22*
6	**Designing your own research**	**23**
	Case study box 10: Comparing SQA results with others	*24*
6.1	Ethics	25
	Case study box 11: Ethics in medical trials	*26*
7	**In conclusion**	**27**
8	**Further reading**	**28**

Introduction

Evidence based decision making is nothing new, it has its origins in the scientific research conducted in agricultural field trials (both crops and livestock) pioneered by the evolutionary biologist and statistician Ronald Fisher at Rothamsted Experimental Station in the 1920s and 1930s.

From there the scientific analysis of the results of experiments and field trials spread to medical research in the clinical trials of drugs and other medical interventions during the 1970s and 1980s, gathering momentum to its application in many aspects of current public health medicine. Increasingly, there is interest in applying these evidence based decision making methodologies to public services more generally including education [1].

This application of the scientific analysis of the results of experiments, field trials and collected data in education is becoming part of public service policy [2]. The challenge of evidence based research in education is that people are the experimental subjects. The design of experiments, trials and data collection must take account of the inherent variability of the experimental subjects to allow a valid and reliable interpretation of the collected evidence. This is where the forensic nature of the scientific approach is essential, without it the results of collected evidence can be worthless. For example see Case Study Box 1: *A badly conducted trial and how it could have been sorted* [3].

[1] Building Evidence into Education, Ben Goldacre, 2013 at http://www.tactyc.org.uk/pdfs/Goldacre-Paper.pdf (accessed May 2016).
[2] Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials, Laura Haynes, Owain Service, Ben Goldacre, David Torgerson, 2012, Cabinet Office Behavioural Insights Team at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf (accessed May 2016).
[3] Fish oil supplements - is there evidence that they improve concentration and behaviour in children? UK Medicines Information (UKMi), 2013, NHS at http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=22&ved=0CCYQFjABOBQ&url=http%3A%2F%2Fwww.medicinesresources.nhs.uk%2FGetDocument.aspx%3FpageId%3D504108&ei=h2znVNSJIliQ7AaGt4G4Dg&usg=AFQjCNFhXuSru17AgKonVJKCB0Xgmsnaow (accessed May 2016).

CASE STUDY BOX 1

A badly conducted trial and how it could have been sorted

Fish oils are a popular food supplement product in the UK. Oily fish are a source of a number of essential nutrients and the UK Government's Scientific Advisory Committee on Nutrition recommends that we should eat at least two portions of fish per week, one of which should be oily. A number of unsubstantiated claims are made for the health benefits of fish oil supplements based on the physiological roles of the nutrients they contain.

Durham County Council set out to give 5000 school children six fish oil capsules per day in their GCSE year and then compare their exams results to their predicted grades. The company supplying the fish oil pills claimed that their formula can help to enhance achievement in the classroom. Press releases and media reports confidently predicted that positive results would be achieved. Ben Goldacre took them to task in Chapter 8 'Pill Solves Complex Social Problem' of his book Bad Science [4].

This was a missed opportunity for a properly conducted randomised controlled trial. A null hypothesis would have been a more scientific starting point than a confident prediction of success. Children taking fish oil capsules should have been compared to a control group. Children should have been allocated to the treatment and control groups randomly. The control group should have been given a placebo to control the placebo and/or Hawthorne effect. The study should have been conducted as a double blind trial. The impact of other concurrent initiatives to improve exam results should have had their impact on the study reduced by blocking. The trial protocol should have been published in advance, including the statistical methods to be used for evaluating the results, and the protocol adhered to as published throughout the trial.

Evidence based decision making is not always welcomed. Often educational practice is driven by the opinions of eminent figures who see evidence based practice as a challenge to their authority. Resistance of this sort was encountered in agriculture and medicine when evidence based practice first appeared. Evidence based practice relies heavily on using scientific methods to collect evidence but is not the only approach. Social research methods can also be used to collect evidence. This can cause tensions; scientists can be suspicious of social research and social researchers can be suspicious of science. This is a false dichotomy; different methods are useful for answering different questions, teachers should be familiar with both.

Where teachers undertake their own research projects it is often referred to as 'action research'. Action research is defined as a process of enquiry conducted by and for the teacher with the aim of changing and improving their practice. Action research has its origins in social research and consequently tends to focus on using social research methods - this need not be the case; a scientific approach can be taken to the collection of evidence.

[4] Bad Science, Ben Goldacre, Fourth Estate, London, 2008.

Action research is often seen as part of the improvement planning process where you reflect on the effectiveness of your own teaching and your students' learning. Then you devise a plan to address the issue of concern, implement the plan, observe, record and analyse the results and reflect on your findings. This may lead to a change in practice (or not) or to further research. Action research forms part of the teaching process; it should not detract from nor disrupt teaching.

The key elements to be considered in action research are:
- Decide on a focus for the research.
- Develop a plan for the research.
- Collect and analyse observations and data.
- Reflect on and evaluate your findings and identify future courses of action.
- Share the findings with others.

Further information on action research can be found in the SSERC publication The Modern Science Teacher [5].

This document describes the main methods for collecting evidence on which to base decisions. These methods are:
- Randomised controlled trials (considered the gold standard of evidence because they demonstrate cause and effect)
- Correlation studies (which show the relationship or link between different factors or conditions)
- Systematic reviews and meta-analyses (used for large scale literature reviews – something to look for in your reading rather than carry out yourself)
- Social research methods (such as questionnaires, interviews and observations).

Teachers can and should employ all of these methods, selecting the ones that are most appropriate for what they want to find out – a case of 'horses for courses'. Research that teachers carry out will inevitably be small scale, but is of no less value for that. The teacher researcher simply has to be careful about generalising their results to a larger population than their research sample and be cautious about the transferability of their results to other situations where circumstances may be different.

This document then goes on to provide advice on designing your own research and analysing results. Finally (In Conclusion) it considers the challenges in getting evidence based practice accepted to become part of everyday practice in education.

Throughout this document the principles of using evidence based practice are illustrated in Case Study Boxes. The document can be read from start to finish, or the main text can be read and then a selection (or all) of the Case Study Boxes can be read. Alternatively the reader can read some (or all) of the Case Study Boxes and then read the main text for further explanation. All of the references selected are freely available electronically; none are behind pay walls so that the reader can readily consult them. There is also a section of further reading at the end. This consists of three relatively short papers; two that make the case for using randomised controlled trials and one that makes the case for a more holistic approach using social research methods. These are well worth reading (it will not take long) for those readers who want to think about evidence based practice more deeply.

[5] The Modern Science Teacher, Jim Stafford, SSERC, 2013 at http://www.sserc.org.uk/images/Publications/MST_book_2013_web.pdf (accessed May 2016).

Randomised controlled trials - the gold standard of evidence

1

In a randomised controlled trial we seek to establish cause and effect. That is we will treat a group of subjects in a particular way or apply an educational intervention to that group. This group becomes the experimental group (or in scientific terms the experimental treatment). This group will then be compared to a control group that does not receive the treatment or intervention (the control treatment). Hence why it is called a controlled trial - we compare the experimental group to a control group.

We also attempt to remove (control) the effect of any other factors (confounding variables) that might affect the outcome of our study. For example if the outcome of our trial was measured in some kind of test we might have half of the test papers in reverse order in case the order in which the questions are attempted influences the result (an order effect).

We can have several different treatments in a trial if we want to compare different approaches or interventions (a multi-factorial trial). We can also have different control groups. In many science experiments the control treatment is the absence of the experimental treatment, in education that is less likely to be the case as if we were trialling a new teaching approach it is unlikely that we would have a control group that received no teaching! Thus the control group is likely to be our existing teaching approach (the control treatment). Control groups that receive no treatment are referred to as negative controls. We can also have a positive control; that is a treatment that shows the desired effect, so that we can see how our new experimental treatment matches up. So a controlled trial could have several different treatment groups and several control groups. There are pros and cons to such large multi-factorial trials compared to a one treatment, one control design; but if you are looking to dip your toe into educational research in a science department it might be best to start out with a more simple trial design comprising of one experimental treatment and a control.

1.1 THE PLACEBO EFFECT, BLIND AND DOUBLE BLIND TRIALS

The placebo effect is a well recognised feature of clinical trials. A placebo is often a pill that is identical in every way to the pill used in the experimental treatment except that is does not contain the active ingredient (it is often replaced with sugar). The placebo is a control. The interesting feature of using a placebo is that, in clinical trials, it gives a better outcome than no treatment at all. It appears that being part of a clinical trial with all its procedures and attention given to patients actually improves medical outcomes. A similar effect occurs in studies that measure the performance of individuals participating in a trial, it is called the Hawthorne effect. Children will do better just from being part of a special group that is being studied, observed and closely attended to. That is if children are told they are part of a special study designed to improve performance, then their performance will improve.

The Hawthorne effect has consequences for the design of any trial we might wish to conduct. If the experimental subjects know they are part of a trial to improve performance then they will perform better than a negative control irrespective of the experimental treatment applied. The way around this is to have a ‘blind’ procedure where the subjects are either unaware that they are part of a trial (which might have ethical implications - we will discuss this later) or are unaware whether they are in the treatment or control group; they are blind to the treatment they are receiving. There is also the possibility that the person conducting the trial will have a preconceived idea about the likely outcome of the trial (in all likelihood they will be testing their own hypothesis.). To avoid any subconscious or deliberate bias in conducting the trial or assessing its outcome, it is best that the experimenter is also blind to which subjects are in the treatment group and which are in the control group. This is usually achieved by coding the individuals in each group. Trials where both the experimental subjects and the experimenter are blind to which subjects are in which group, are referred to as double blind. Wherever possible, trials with human subjects should be double blind.

1.2 HISTORICAL AND OTHER SIMILAR TYPES OF CONTROL

A historical control is one where we compare the results of our treatment group with the results of previous studies. In general this is best avoided and you are best to use a concurrent control built into your own study. The only difference between a historical control group and a current treatment group should be the different treatments in the two groups; all the other experimental conditions should be the same and that can be difficult to guarantee if you are using someone else's results or results gathered previously. This emphasises the importance of the methodology, conditions and participants being described accurately so that another worker can repeat the study precisely. However on occasions it may be necessary to draw historical comparisons. For example the SQA from time to time monitors the maintenance of examination standards by comparing performance in current examinations with previous ones. Clearly we cannot compare entire exam papers from different time periods; rather the researchers have got to look for questions on the same topic at a similar standard with equivalent mark schemes from both papers and then compare the scores of candidates who have achieved similar awards in the two examinations (see also Case Study Box 2 *Can we compare the results of National 5, Intermediate 2 and Standard Grade?*).


Image: Kenn W. Kiser/www.morguefile.com

CASE STUDY BOX 2

Can we compare the results of National 5, Intermediate 2 and Standard Grade?

Much was made in the press at the time of the first National awards that the results of Standard Grade and National 4 and 5 could not be compared. Although it is true that these are different awards and cannot be compared in an overall way, the qualifications do contain benchmarked standards that allow for comparison. National 5 is benchmarked to credit at Standard Grade which in turn is bench marked to Intermediate 2. Thus awards at each of those levels can be compared. In the school session 2014/15 there was a S4 cohort where presentations were made in both Intermediate 2 and National 5. This would allow comparisons to be made between these qualifications without having to use a historical comparison. In turn in the previous session presentations were made at both Intermediate 2 and Standard Grade allowing comparisons to be made between these two awards. Thus Intermediate 2 could be used as an intermediary (no pun intended!) to compare Standard Grade and National 5. Equally a direct historical comparison could be made between these qualifications using different cohorts. In making such comparisons it is important to use equivalent samples for each qualification. Making use of a mean exclusive points (MEP) score for all a candidate's subjects could be used as a basis for creating samples that are equivalent in ability. The mean exclusive points (MEP) score is defined as the mean tariff points [6] for all the candidate's courses.

A more general concern that has been raised on the replacement of Standard Grade with National Qualifications is that a reduction in the number of examination subjects taken in S4 will lead to a fall in the number of learners taking particular subjects. In particular will the fall in numbers in subjects be proportionate to the overall fall in the number of subjects taken or will some subjects experience a greater reduction in numbers than others? This may be exacerbated by the removal of the requirement for learners to study each of the Standard Grade modes; although that restriction may continue to apply depending on the curriculum course choice options offered. Tracking such changes in subject uptake will require taking these considerations into account. One simple way of tracking subject uptake is to measure the 'science factor'. Simply add the numbers of learners taking each science subject and divide it by the number of learners in the year group. Previous calculations in the Standard Grade curriculum over a range of schools showed an average science factor of 1.27. Another potential impact of the reduction in the number of subjects taken in S4 is the effect on subject choice in S5. Previous studies have shown that a significant number of learners when choosing S5 subjects chose subjects from outwith the subjects they selected at Standard Grade. To what extent will the later choice of S4 subjects, the reduced number of subjects taken and curriculum choice options offered in S4 have on learners continuing their subjects of study in S4 into S5?

We also have to exercise caution where a controlled trial takes place over an extended period of time. For example, imagine we introduced an initiative in S1 with a view to improving exam performance in S4. In the intervening period of time the people in each group will change, some will mature and get smarter; others will lose interest and motivation and these changes might be responsible for the changes we observe. This is no different to clinical trials where the results of a

[6] Tariff scale applied in Insight, Scottish Government: http://insight-guides.scotxed.net/support/InsightTariff.pdf (accessed May 2016).

drug trial may be confounded by individuals who get better of their own accord (for example due to their immune response) rather than due to the drug intervention (good for them, but it does mess up the experimental design!). Case Study Box 3 *Base line studies*, considers trials that occur over a time line and historical controls.

CASE STUDY BOX 3

Base line studies

There are three main types of base line study. One form of base line study is where we measure something of interest, apply an intervention and then apply the measurement again to see if the intervention has had an effect. In this case it is important to have a control group which does not receive the intervention so that we can assess the impact of the intervention. Subjects should be allocated at random to the two groups and the groups checked for in group variation and to see if a form of blocking is required. Studies of this type can be used to assess teaching strategies to resolve misconceptions. Test both groups to see the strength of the misconception, apply the treatment to address the misconception to one group and then reassess each group to see the impact of intervention. A further assessment after an extended period of time might be of interest to see if the subjects revert to the misconception. Because misconceptions are often based on logical constructs they can be persistent and reappear after the passage of time.

Baseline studies can also be used to monitor performance, for example in year on year comparisons of a department's exam performance. Such comparisons can be of value but only if it is made on a like for like basis. Here it is important to compare the nature of successive year cohorts to make fair comparisons. A measure of cohort overall ability would be important for example. Other baseline studies of this type could examine gender balance in successive subject cohorts or the uptake of individual science subjects and combinations of science subjects. This kind of data could be a regular feature of a department's own internal annual review or Standards and Quality Report.

Another type of baseline study is where use is made of historical data held on a group of individuals to study the impact of time and any intervening circumstances since the original data was recorded. One of the best known studies of this type is the 'Lothian Birth Cohort Study'. In 1932 and 1947 almost all the children in Scotland born in 1921 and 1936 (11 years old) undertook an intelligence test. Samples of the people who had taken part in these studies have been traced, recruited and retested along with surveys of their lifestyle factors and medical histories in a variety of research studies. For example, 1080 men and women from the 1936 Lothian Birth Cohort Study were retested at age 70 and surveyed in relation to their smoking habits [7]. The results were corrected to control the effect of socio-economic status (blocking). Current smokers scored lower than ex smokers and never smokers in general cognitive ability and processing speed, but not for memory or verbal ability.

[7] Smoking, childhood IQ, and cognitive function in old age, Corley J, et al. Journal of Psychosomatic Research, 73, 132-138 (2012) at http://www.research.ed.ac.uk/portal/files/16310520/Corley_2012_Smoking_childhood_IQ_and_cognitive_function.pdf (accessed May 2016).

1.3 RANDOMISATION

When we allocate subjects to treatment and control groups, we should do so randomly to avoid any bias in the groups. Random allocation is different to a haphazard allocation; we should draw lots, use random number tables or use the random number generators that are available in computer software and smart phone apps. Even although the groups will now be equivalent, there may still be considerable random variation between the individuals in each group. If there is a wide variation between the individuals in each group then there will be an increased chance that the differences between the groups may be due to chance rather than the treatment effect or that the variation within the trial groups will mask any treatment effect. Although the statistical analysis of the results of the trial can address these problems we should do what we can to reduce the impact of within group variation on our trial results.

The first and most obvious thing to do is to examine the variation between the individuals in each group and in our study as a whole. Surprisingly, this important initial step is one that is often omitted from studies once a random allocation of individuals to groups is made. The variation in each group should be representative of the variation in the entire sample used in the trial, which should, ideally, in turn be representative of the whole population. Obviously this last stage is difficult to achieve in a single school as the school population may not be representative of the whole Scottish school population. One of the things we can do to address the problem of in group variation is to increase the size of the trial sample and consequently of the treatment and control groups. In doing so it is important to keep the group sizes the same (particularly if we are considering a statistical treatment of our results). Increasing sample and group size increases the likelihood of the within group population being representative of the variation in the whole population.

1.3.1 Blocking

Obviously, increasing the trial sample and group sizes may not always be practicable in a school setting. Another way of addressing the problem of the variation in groups is the technique known as ‘blocking’. Here we look at the factors that vary in our groups and identify any that we think may influence our results. For example if we thought gender was likely to influence our results we would divide our subjects into two blocks, boys and girls, and allocate the treatments randomly within the two blocks. If we thought mathematical ability could be a confounding factor in our trial, then we would rank the individuals by mathematical ability and then divide them into blocks of similar ability. The number in each block should be a multiple of the number of treatment groups. Then again we would allocate the treatments randomly within each block. Sometimes the blocks present themselves to us. Say we thought the teacher is likely to be a factor that will cause variation in our results; then each class would become a block. Such designs are called randomised block designs. Remember that blocking should only be used where, in our sample, we have wide variation of a factor that is likely to influence our results.

Image: cameio/www.morguefile.com

1.3.2 Paired designs

Paired designs are a common form of blocking. Here we divided the population into pairs of similar individuals (or groups) and randomly assign the individuals in each pair to one of two treatments. For example, a local authority wanted to evaluate the impact of the Cognitive Acceleration through Science Education (CASE) Programme [8]. The 32 secondary schools were arranged in pairs with similar exam performance data and one in each pair was randomly assigned to undertake the CASE programme in S1/S2 while the other did not. Once the sample cohorts had worked their way through to S4, the Standard Grade results of the treatment and control groups were compared.

1.3.3 Within subject designs

Another way to reduce the effect of variation in our groups is to use a 'within subject design'. Here the subjects experience the different treatments sequentially, and comparisons are made on the same individual at different times, rather than between different individuals at the same time (a between subject design). This also has the advantage of effectively increasing the number of experimental units, as each subject experiences all the experimental treatments making it a useful technique where we have small samples. Although a within subject design removes many of the problems associated with variation between the subjects in a sample or experimental group, it is not suitable for situations where the subjects cannot be returned to the condition they were in before a treatment or intervention was applied (reversibility). For example, if we wanted to trial two different strategies for teaching molarity calculations in chemistry it wouldn't make sense to do that sequentially as there is every likelihood that there would be a 'carry over' effect from the first treatment that would influence the second treatment. Where we are using a within subject design we should 'counter balance' our treatments to eliminate any order effects. In a trial with two treatments, half the subjects would experience treatment 1 followed by treatment 2 and the other half treatment 2 followed by treatment 1. It is a good idea to leave a time interval (a wash out period) between the two treatments to help minimise carry over effects. For example you might be interested to compare the accuracy with which students can identify invertebrates (mini beasts) in leaf litter using either line drawings or photographs of the invertebrates commonly found in leaf litter. Half of the subjects would start with line drawings followed by photographs and the other half the reverse. You would allow a suitable time interval between the two treatments and not provide the subjects with any feedback until both treatments were complete.

Image on left by Ian Alexander (Own work) [CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0)], via Wikimedia Commons.

[8] The science of thinking, and science for thinking: a description of cognitive acceleration through science education (CASE), Philip Adey, International Bureau of Education, UNESCO, Geneva, 1999 at http://www.ibe.unesco.org/publications/innodata/inno02.pdf (accessed May 2016).

Correlation studies 2

In a correlation study we measure naturally occurring factors (variables) and draw comparisons between them rather than compare experimental treatments to controls as we do in a randomised controlled trial. This has the attraction of using naturally occurring evidence rather than the manipulation of a randomised controlled trial.

Correlation studies have the advantage of avoiding any unintended effects that may arise as a result of the manipulations in a randomised controlled trial and of the subjects being aware that they are part of a trial. However, correlation studies have two significant potential draw backs. Because we have not controlled the experimental conditions, we do not know if there may be other variables that are influencing our observed measurements. For example when the number of ice cream sales rise, so does the number of drowning incidents. This is a positive correlation or association between these two measured observations (as one variable rises so does the other). It is more likely that this is due to a third variable, warm weather, during which both of these conditions increase rather than any direct connection between eating ice cream and drowning. Also when looking for an association between observations we do not know which variable influences which. We may mistakenly assume that factor A influences factor B when in fact it is factor B that is responsible for the change in A; a situation we refer to as 'reverse causation'. For example a study showed that children who were put to sleep with the light on had a higher incidence of myopia (short-sightedness) [9]; suggesting that light while sleeping is responsible for myopia. However further studies [10] showed that parents who suffered from myopia were more likely to leave their child's light on (we can imagine why). Thus it appears more likely that myopia causes people to keep the light on, rather than keeping the light on causes myopia.

All of this demonstrates an important distinction between randomised controlled trials and correlation or association studies. A properly conducted randomised controlled trial demonstrates cause and effect; a correlation study can only show that there is an association (positive or negative) between the observations. We can sum this up by saying that in a randomised control trial we can draw conclusions beyond reasonable doubt but in a correlation study our conclusions are based on the balance of probabilities. However we should not be too harsh in our judgement of correlation studies, often they can provide us with the basis of a hypothesis that can then be tested in a randomised control trial. For example see Case Study Box 4 *Can drinking water improve exam performance?* One of the best known correlation studies is the British Doctors Smoking Study conducted from the 1950s to 2001. Data was collected from a survey of 35,000 British doctors that recorded their age, smoking habits, physical health and cause of death. The data showed a positive correlation between the incidence of lung cancer and continued smoking and the daily number of cigarettes smoked. The results from controlled laboratory experiments with animals were then used to establish the causative link between smoking and lung cancer beyond doubt.

[9] Night-light may lead to nearsightedness, CNN, 1999 at http://edition.cnn.com/HEALTH/9905/12/children.lights/index.html (accessed May 2016),

[10] Night lights don't lead to nearsightedness, study suggests, Research News, The Ohio State University, 2000 at http://researchnews.osu.edu/archive/nitelite.htm (accessed May 2016).

CASE STUDY BOX 4

Can drinking water improve exam performance?

In a university study of 447 students, exam performance was related to whether students brought water into the exam [11]. The study used coursework marks to control for any difference in ability between those students who brought water into the exam and those that did not. The study showed a 5% improvement in marks for students who brought water into the exam. The researchers raised the possibility that water consumption may have a physiological effect on thinking functions that influence exam performance. They also raised the possibility that drinking water may alleviate the anxiety that can influence exam performance.

This is a correlation study. The group of students who brought water into the exam were self selecting rather than being allocated at random as would have been the case in a randomised controlled trial. It might be the case that the group of students who bring water into the exam have other properties that were not controlled in the study. For example they may have a more organised preparation regime for exams than those students who do not bring water into the exam. This is an opportunity for a randomised controlled trial using a within subjects design. For example, a group of students could be divided randomly into two groups. Both groups are given a test, with only one group being supplied with drinking water and the test results of both groups recorded. Then after a suitable time interval, another similar test is applied to the two groups with the supply of water reversed between the two groups. This design should 'counter balance' any order effect in the treatments (whether drinking water is made available in the first or second test). Using a within subjects design effectively increases the number of experimental subjects (each student is the subject of both treatments) and reduces the differences between treatment groups (all students experience both treatments). To analyse the collected results we group all the results into those where drinking water was supplied and those where it was not to make our comparisons.

There is also an opportunity here to investigate a placebo effect. We could compare taking drinking water into an exam with other observed student practices. For example comparing drinking water during exams with chewing gum or eating polo mints. In any event it makes sense to be properly hydrated and relaxed for exams!

[11] Can water boost your exam grades? The British Psychological Society, 2012 at http://www.bps.org.uk/news/can-water-boost-your-exam-grades (accessed May 2016).

2.1 CASE CONTROL STUDIES

On occasion you might come across data on an outcome of interest and wish that you had done a correlation study or perhaps a randomised controlled trial. All is not lost, as you can still identify a suitable control group and carry out a retrospective study (had you designed a correlation study or a randomised controlled trial in advance that would be referred to as a prospective study). This is called a 'case control study' as we start with a group of individuals that show the outcome of interest (the case group). For example, in Australia the benefits of a school based bicycle education programme were evaluated using a case control study [12]. The 'case group' were children presenting at hospital A&E departments with injuries received while riding bicycles. The control group was recruited by dialling randomly selected telephone numbers. Data were collected by personal interview with all subjects. An analysis of the results concluded that the bicycle safety education programme does not reduce the incidence of bicycle injury in children. See also Case Study Box 5 *Are Easter school revision classes and study clubs any good?*

CASE STUDY BOX 5

Are Easter school revision classes and study clubs any good?

Schools and other providers often organise after school study clubs and Easter school revision classes during the school holidays prior to SQA examinations. We might wonder if they do any good and how would we know? The obvious thing to do might be to run a randomised controlled trial comparing a group that attends such provision and a group that does not. This could be difficult to set up. It is unlikely that you would be able to allocate individuals to each group at random. Attendance is likely to be voluntary (with implicit consent) and therefore self selecting. Some pupils may be desperate to attend and might demand to be part of the 'treatment' group rather than the control group and for others attending additional classes in their own time would be the last thing they would want to do. One option could be to conduct a case control study. We could look at the attainment of a group of learners that have participated in such provision and then seek out an equivalent group of learners that did not and compare the attainment of the two groups. Remember that such a study would be a correlation rather than a randomised controlled trial, so we would need to be cautious in drawing conclusions from the results.

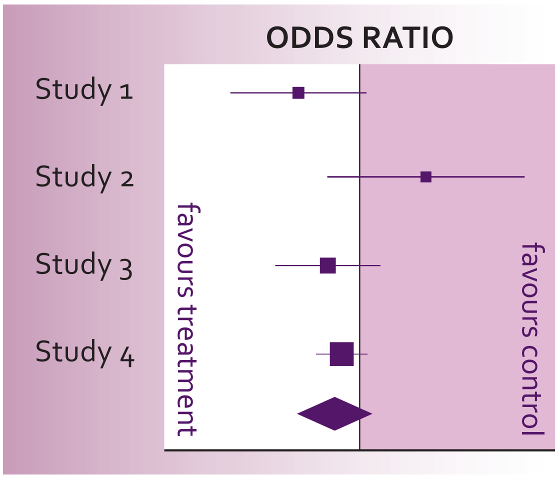
We might want to select individuals that are 'borderline' candidates for a pass for both groups. If we were looking at revision classes for Higher then we would want to look at information that is the best predictor of Higher performance. From research in the past we know that the 'grade point average' for all a candidate's Standard Grade subjects is a better predictor of future Higher success than their performance in an individual subject. In particular we know that candidates with a grade point average of 2.0 tend to be borderline pass candidates at Higher. We would need to determine what a Standard Grade grade point average of 2.0 equates to at National 5. That should not be too difficult (although some refinement might be necessary) as Standard Grade credit is benchmarked at National 5. Then we could select a sample of predicted borderline pass candidates from the study club/revision class population and select an equivalent borderline pass sample from candidates who did not attend a study club/revision class. This is a case control study using a matched sample blocking design.

[12] School based bicycle safety education and bicycle injuries in children: a case-control study, Carlin JB, Taylor P, Nolan T, Injury Prevention, 1998 at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1730310/pdf/v004p00022.pdf (accessed May 2016).

3 Systematic reviews and meta-analyses

A systematic review is a survey of all the relevant research data that can be found on a topic without bias towards any particular set of findings or research methodology. The researcher specifies their search criteria in detail and presents everything they have found including any research rejected with an explanation of why. The next stage is the meta-analysis of all the findings.

The treatment effect is calculated (often as an odds ratio) for each study along with the variation (confidence interval) in the results. In the next stage an overall treatment effect is calculated as a weighted average of the individual summary statistics. Greater weight is given to the results from studies that provide more information.



The results can be shown in a 'blobbogram' or 'forest plot' as shown in the diagram below. The central vertical line is the 'line of no effect'. The trials with longer horizontal lines show more variation in their results (often due to small sample size) and where they touch or cross the middle vertical line the results are not statistically significant. The squares show the treatment effect for each trial. Trials with larger squares have been given greater weighting. In the diagram, study 4 has been given the greatest weighting and shows the smallest variation so it is likely that it is a large well conducted study. The diamond shape shows the overall treatment effect with the centre of the diamond showing the combined effect and the tips of the diamond show the overall variation in the results. Notice that in this case the overall positive treatment effect is not statistically significant as the diamond touches the central line.

The best known examples of systematic review and meta-analysis in school education are Hattie's work on 'Influences on Student Learning' and 'Teachers make a Difference' and the work of Joyce and Showers on 'Student Achievement through Staff Development'. Although these researchers have published both books and research papers on their work it is in the nature of systematic review and meta-analysis to continually update findings as new work becomes published. These workers have both done this and continued with their own research. As a consequence rather than quote references here it is safer to look via search engines to obtain a current picture of this research.

In his work on Influences on Student Learning, Hattie now (2014) has a list of 138 influences graded by their effect size. These include some recent additions including Piagetian programmes (of which CASE can be considered one). Some established practices such as homework and ability grouping are shown to have a limited positive effect compared to others. In interpreting this research caution is required to make sure you have a clear understanding of what Hattie means by each of his influences as this is not always clear from looking at a table ranking these influences. However his research supports Piagetian programmes such as CASE and has been a factor in the Assesment is for Learning (AiFL) movement which promotes comment only marking (for example two stars and a wish) rather than scores and grades (Hattie's feedback influence). Also the use of

Concept Cartoons [13] in science learning can be considered a form of reciprocal learning which also scores highly in Hattie's effect size.

In his work on Teachers Make a Difference Hattie makes a distinction between novice, experienced and expert teachers. In his work on Influences on Student Learning he categorises the influences as being the students themselves, teachers, the school, peers and home. Of these the student is the greatest influence (about 50%) followed by teachers (about 30%). Hattie is at pains to point out his distinction is not between good and bad teachers; all teachers will make the journey from novice to experienced as they learn their craft but not all will become expert teachers. Students taught by expert teachers exhibit an understanding of concepts that is more integrated, more coherent, and at a higher level of abstraction than the understanding achieved by other students. The elements that best distinguish expert from experienced teachers are: challenge, deep representation and monitoring and feedback. In science, engaging students in challenging activities will include evaluating scientific work of self and others, being creative in designing experiments and investigations and applying knowledge and skills to less familiar and more complex situations including real life situations. Deep representation is about deeper, more detailed and better connected learning about science. It involves relating and extending ideas to develop understanding rather than the surface learning of acquiring knowledge. Expert teachers can anticipate and prevent difficulties in science learning rather than correct existing difficulties; they monitor and get feedback from learners to gain insight in how to develop and test strategies for learning.

Joyce and Showers' initial research used systematic review to examine the effectiveness of different methods of teacher in-service training on changing teachers' classroom practice. They had previously observed that in-service teacher training often did not result in the desired change becoming embedded in teachers' practice. They examined five components of teacher training and found the following:

- Imparting new knowledge by lectures and information giving sessions had little effect on changing teachers' practice.
- Observing demonstration lessons or someone modelling a new skill or strategy improves the teacher's understanding of the new method, but alone has little impact on changing practice.
- Practice of a new skill or strategy in a simulated or protected setting develops competence and confidence in the new technique. Although a good proportion of teachers will then alter their practice, without continued support many will revert to their existing practice.
- Feedback based on observation of the new teaching skill or strategy helped to embed the skill and increased the proportion of teachers who embedded the change in their practice.
- Coaching by peers on how to apply the new skills and strategies, successfully applied, brings about the embedded change desired in almost all cases.

The insight of Joyce and Showers was to appreciate that these methods should not be used in isolation but in combination and indeed in sequence. Criticism is often levelled at information giving sessions being of little value, but the evidence shows that is not the case; rather it is the lack of following up with the rest of the sequence that result in poor outcomes. Joyce and Showers then went on to trial the sequence with and without feedback sessions and as a result changed to a four stage model omitting feedback. It is interesting to compare this with Hattie's work with students where feedback is very effective, compared to teachers where peer coaching is more successful. More information on Joyce and Showers research and its implications can be found in Chapter 6 Leading Professional Learning in the SSERC publication The Excellent Science Department [14].

[13] Concept Cartoons: What Have We Learnt?, Stuart Naylor and Brenda Keogh, Journal of Turkish Science Education, 2013 at http://www.tused.org/internet/tused/archive/v10/i1/tusedv10i1s1.pdf (accessed May 2016).
[14] Jim Stafford, The Excellent Science Department, SSERC, 2014 at http://www.sserc.org.uk/images/Leadership/ESD_book%202014%20web.pdf (accessed May 2016).

Social research methods

4

Social research often involves the collection of information through methods such as questionnaires, interviews or observations. The information from social research can be either quantitative or qualitative. The results of quantitative research methods can be recorded as counts in various categories to give numerical data that can be analysed; qualitative research is often recorded as a descriptive narrative describing the responses.

Some see a tension between such qualitative research and randomised controlled trials. This is not so. Randomised controlled trials are good at showing that something works but not necessarily why it works. Qualitative research can help to provide an insight into how something works. Qualitative research can also generate ideas for hypotheses that can then be tested quantitatively using randomised controlled trials or correlation studies. The secret is to employ the right methods to answer the right questions. Often a variety of approaches can be used to complement each other in a research study. For example see Wynne Harlen [15].

Social research is usually relatively small scale and the researcher is intimately involved in the collection and analysis of the findings. In social research the researcher's identity, values and beliefs cannot be entirely eliminated from having an influence on the research. The results of social research can be subject to unintentional bias due to the selection of the participants or due to the preconceptions of the researcher. Thus the researcher has to consider how the potential for prejudice and bias can be avoided to make the collection and analysis of the data fair and even-handed. This is a question of reliability; in other words would the research produce the same results in the hands of another researcher? There are a number of techniques that can be used to address this reliability issue. One is the process of 'triangulation' where evidence is compared from different perspectives. The perspectives triangulated could be from different sources of information (for example from documents, interviews and observations) or from different researchers. For example see Case Study Box 6 *How school inspections work*.

The small scale of many social research projects may not a problem if the results are only going to be applied locally (as is the case in action research). However the small sample size often means the results cannot be applied more generally and transferred to other situations. As a consequence it is useful if an audit trail of how the evidence was collected is presented along with the research findings including information on the nature of the participants involved in the research. This information can help to make the findings more transferrable to other situations.

Image: Daniela Goulart/www.flickr.com

[15] Reflections on a personal journey in research, Wynne Harlen, ASE, 2009 at http://www.ase.org.uk/resources/scitutors/research/research-wynne-harlen/ (accessed May 2016).

CASE STUDY BOX 6

How school inspections work

School inspections employ social research methods. During a school inspection, inspectors gather evidence from a variety of sources. They examine documentary information such as school attainment data, the school's self evaluation report, standards and quality report, improvement plan, tracking achievement information, pupil progress records, course choice information, staff handbook, record of professional learning. Inspectors gather information on learning and teaching through their observation of class lessons. They collect evidence of the views of pupils, parents and staff through interviews, focus groups and questionnaires. The evidence from documents, observations and interviews/questionnaires is then triangulated for the purposes of evaluating and reporting.

A detailed (and confidential) record is kept of the findings in the inspection, the Record of Inspection Findings (RIF). This document acts as an audit trail of the collected evidence and is made available to the school. When observing lessons, inspectors record their findings using a lesson observation record which acts like an observation schedule. The lesson observation record is based on a selection of the Quality Indicators from How Good is our School? (*HGIOS?*).

As a result of the inspection, inspectors produce a letter for parents which provides a narrative report on how well young people are learning and achieving, are supported by the school in their learning and on the school's ability to improve. In addition the school is sent information on the evaluation of Quality Indicators on a six point scale and the analyses of questionnaire returns. These are published on-line. The six point scale used for the evaluation of Quality Indicators is described as a narrative for each level. There is a general narrative describing the standard for each level and for each Quality Indicator there is a narrative illustration of the standard for level five. If you want to know more about HMI expectations and the day to day pattern of inspection activities then you should consult the *Education Scotland Inspection Advice Note* and the *Briefing note for head teachers of secondary schools*. Both of these are updated for each school session, so rather than give references here it is better to put these into a search engine along with the date of the current session.

4.1 QUESTIONNAIRES

Questionnaires can use either open or closed questions. Closed questions require the respondent to choose from a range of supplied responses. The responses to a closed question can be restricted to as few as two, for example yes/no, true/false, or agree/disagree, or a more extended response scale can be used. Generally five is the best number of points to have on a more extended response scale as longer scales give less consistent results; for example the five-point Likert Scale of strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. The five-point Likert Scale can be adapted by replacing agree/disagree with important/unimportant, difficult/easy, boring/interesting etc as appropriate. Open questions leave the respondent to decide the wording, length and nature of the response. The questions tend to be short and the answers tend to be long; for example "What do you think of ..., What do you like about ... Closed questions lend themselves to data that can be quantified and compared. However the responses may not allow respondents the opportunity to express their views. Although open questions allow respondents to express themselves, the responses can be difficult to analyse and the results may need to be reported as a narrative rather than as quantitative data. Closed questions are less demanding to complete than open questions and so may result in more people completing questionnaires. It is important that the questions are on topics where the respondents have the necessary background knowledge, experience or opinions to provide answers. Like exam questions, make sure the wording of questions is short, straightforward and unambiguous, uses appropriate wording for the respondents and avoids 'leading questions'. It can sometimes be useful to combine closed and open questions in the one questionnaire by asking respondents to explain or comment on why they have chosen a particular option in a closed question. Then when reporting the proportions for each option it can be followed by a narrative describing the responses to the subsequent open question which may give insight into the thinking behind respondents' choices.

4.2 INTERVIEWS

The purpose of a research interview is to gain an in-depth insight into the thoughts, opinions, and emotions of those being interviewed. Although the researcher will have a clear list of issues to be addressed and questions to be answered, the researcher's role is to be as unintrusive as possible and to allow the interviewee to develop their own thoughts and to express themselves. Interviews can be conducted on a one-to-one basis or as a focus group. A focus group consists typically of six to nine people where the researcher adopts a neutral non-committal stance to allow the group members to interact without being led by the researcher. This can be a challenge where the interviewees know the researcher as their teacher. Thus the nature of the interview or focus group and the role of the 'researcher' within it will need to be explained to participants. Emphasise the importance of honesty rather than trying to "help" the interviewer by giving especially clever or imaginative answers. Emphasise the anonymity with which answers will be treated.

Successful interviewing requires the same skills as in a counselling relationship:
- being non-judgemental
- suspending personal opinions and values
- avoiding passing judgement on comments made.

Make use of listening skills - a good guide is that the researcher should talk for no more than 20% of the time:
- Use non-verbal cues to demonstrate you are paying attention to what is being said.
- Use open-ended questions that allow the interviewee(s) the opportunity to speak and elaborate on what they have said.
- Tolerate periods of silence.
- Use prompts such as repeating or paraphrasing what has been said to encourage interviewees to elaborate on what they have said.
- Use probing questions such as asking for examples or more details to gain clarification.
- Summarise what has been said to check with the interviewee(s) you have correctly interpreted what has been said (this will also help you when recording the outcome of the interview).
- In focus groups give everyone the opportunity to have a say and avoid dominant personalities hogging the discussion.

In short the researcher facilitates the interview or group discussion rather than leads it allowing the respondent(s) to speak freely.

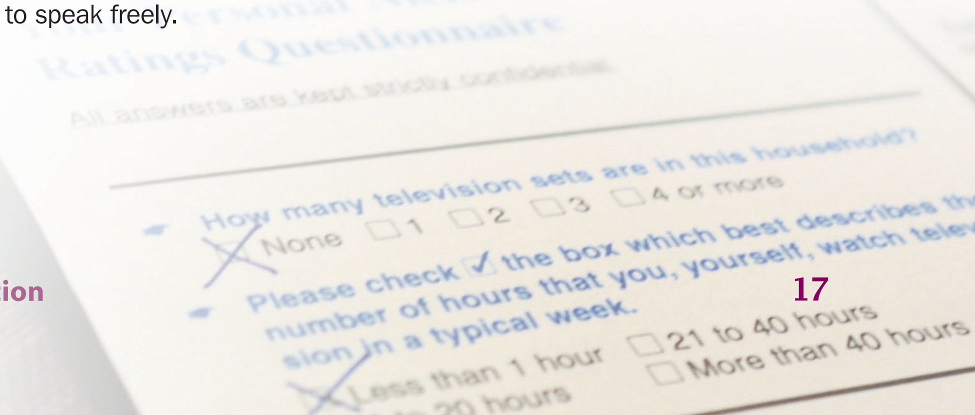


Image: Joe Gratz/www.flickr.com

4.3 OBSERVATION

Basing research on observation has to deal with the potential differences in the collection and interpretation of the observed evidence by different researchers. What a researcher observes will be influenced by selective memory, selective perception, past experience and their emotional state. The way to minimise, possibly eliminate, the variations that arise from different individual perceptions of events and situations is through systematic observations using an observation schedule. In principle this is similar to the ethograms used in animal behaviour studies. The observation schedule works something like a checklist which contains the behaviours or events that are to be recorded.

Selecting the behaviours or events to include in the observation schedule requires some thought. A preliminary study to identify all the behaviours observed in the group for study will help you to select the most significant and the most relevant items to the aims of the observation study. The behaviours should be distinct and directly observable, not requiring any inference on the part of the researcher. Reviewing previous research or using your past experience may also help you to select items for the observation schedule. Then once an event occurs it becomes a matter of recording the frequency or the duration of the event. Behaviours that are short and distinctive are best recorded as the number of occurrences in unit time (the frequency), behaviours that last for longer periods are best recorded as the length of time the behaviour lasts (the duration). Measuring frequency and duration together provides two complementary methods of describing the observed behaviour. Once you have selected the behaviours for your observation schedule, observing and recording them continuously for an entire group can prove challenging. One way around this is to use sampling. You can either record your observations with a selected subset of the larger group or use time sampling where at specific points in time you record the occurrence of the behaviour in the whole group.

Analysing results

5

Be wary about jumping to conclusions about your results or the results of others. Be sceptical but do not be cynical. Scepticism comes from the scientific premise that our understanding is based on the best interpretation we can make of present evidence and knowledge and may change in the light of new knowledge and understanding. Cynicism is often about denying evidence that does not suit our preconceived ideas or purpose. Scepticism leads to the truth, cynicism leads to the results you wish to see.

The first consideration in analysing results should be to examine their validity and reliability.

Validity means that the results provide the answer to the question the research set out to answer. In scientific terms does the experimental design test the aims and/or hypothesis of the study? Remember results that show no effect can be valid. Questions to ask about validity include:

- Are the treatments/variables measured relevant to the aims of the study/appropriate for the hypothesis?
- Are there appropriate controls to show effects?
- Are confounding factors controlled or measured?

Reliability means that the results can be trusted. Questions to ask about reliability include:

- Are samples of sufficient size?
- Are samples representative of the population?
- Are samples randomised?
- Has the variation within samples been examined?
- Have measures been taken to remove observer bias and participant effects? (e.g. blind/double blind trials, placebo/Hawthorne effects and order effects)

Use tables, graphs and charts to summarise results and show trends. Where appropriate use tests of statistical significance built into the design of trials and studies. For example see Case Study Box 7 *Using Statistics to Analyse Results*. Don't be blinded by data and statistics; use your knowledge and experience when interpreting results. For example see Case Study Box 8 *Should Primary Teachers have a pass in Higher Mathematics?* Make the distinction between change and improvement. Do the results show a method that is better, just the same or worse? See Case Study Box 9 *Change for Change's sake?*

Image: StuartMiles/FreeDigitalPhotos.net

CASE STUDY BOX 7

Using statistics to analyse results

Using statistical tests is not as daunting as many teachers fear. The key is to understand the principles of what the statistics can do and apply the correct test for the design of your particular study. In the majority of cases comparisons will be drawn between groups (samples) of learners. Often comparisons between groups are made on the basis of an average or mean value. The mean value does not take into account the spread or dispersion of the values within each group and may lead to dubious conclusions. Statistics makes a comparison between groups on the basis of the distribution of values in the groups rather a single mean value.

The dispersion of values in a sample can be visualised by a box plot. Box plots show the range of values in the sample, the inter-quartile range (the range of the middle 50% of the values) and the median (middle value). This information can also be described by the 'five figure summary' – the minimum, lower quartile, median, upper quartile and maximum values. Box plots of different groups can be presented side by side for comparison. The median is often a better central measure of the values in a sample than the mean as it is less prone to distortion by extreme values.

Statistical tests test for a null hypothesis. That is they test for the assumption that there is no difference between samples. Statistical tests yield a 'p value'. If the p value is less than 0.05 then the chances that the samples are the same is less than one in twenty and they can be said to be significantly different. A good general test that can be used to compare two samples is the Wilcoxon rank sum test (sometimes called the Mann-Whitney U-test). Its advantage is that unlike the t-test it is also suitable for small samples that are not symmetrically distributed. If you are using a within subjects design you should use the Wilcoxon signed rank test. If you have more than two groups or samples, the Kruskal Wallis test is the one to use. The statistical package R is widely used in research and academia and is free to download and will do all of these tests, calculate a five figure summary and draw box plots. Alternatively you can enter the name of the test followed by calculator into a search engine. Further information on statistical analysis and data presentation can be found in the SSERC booklet Statistics for School Biology Experiments and Advanced Higher Projects [16].

[16] Statistics for School Biology Experiments and Advanced Higher Projects, Graeme D Ruxton and Jim Stafford, SSERC, 2015 at http://www.sserc.org.uk/images/Biology/Higher_Biol/Statistics%20book%20final.pdf (accessed May 2016).

CASE STUDY BOX 8

Should primary teachers have a pass in Higher Mathematics?

Consideration is being given to making it a requirement that students entering primary teacher training should have a pass in Higher mathematics. An Initial Teacher Education Institution tested the mathematical knowledge required for primary teaching of its 149 entrants and compared their results to their mathematics qualifications [17]. They found that those with a pass in Higher mathematics performed better than those with a Standard Grade credit award who in turn performed better than those with an Intermediate 2 award. The mean score of students with a Higher was not significantly different from those with a credit Standard Grade but both had significantly greater mean scores than those students with an Intermediate 2 award.

A more meaningful interpretation of this information can be made by making reference to existing knowledge and experience. All the students will have covered the mathematical knowledge necessary to teach primary mathematics in their school qualifications. Experience suggests that a number of the students with an Intermediate 2 award will have first achieved a Standard Grade award at general and are therefore likely not to score as highly as those students who achieved credit Standard Grade at the first attempt. Of those who achieve credit at Standard Grade some may elect not to progress to Higher although they have the ability to do so. This may explain why although those with a Higher have the highest scores their scores are not significantly different from those with credit at Standard Grade. Another consideration is the work of John Hattie on ‘Teachers Make a Difference’. It is reasonable to expect that all students will make the journey from novice to experienced teacher but it may be the case that only the more mathematically able will move from experienced to expert in the realm of mathematics teaching. All of this evidence should be used to inform the decision about entrance qualifications for primary teachers.

Image: Jeremy Mikkola/www.flickr.com

[17] McKechan, S. and Day, S., Do advanced qualifications equate to better mathematical knowledge for primary teaching? - Paper 12 Scottish Educational Research Association, Edinburgh, 2014 (abstract) at http://www.sera.ac.uk/documents/2014/SERA_2014_Book_of_abstracts.pdf (accessed May 2016).

CASE STUDY BOX 9

Change for change's sake?

From time to time educational change is driven by theory and philosophy. The role of research is to formulate questions and/or hypotheses that can test the veracity of such theories and philosophy or to collect observational evidence of their impact. For example some considerable time ago there was a movement to replace traditional didactic approaches in secondary school science education with what was known as Resource Based Learning (RBL). A chemistry department switched their learning and teaching to RBL and monitored its impact on S4 exam results compared to when they were using their more traditional didactic approach. They found their results were much the same. However they did reckon that some pupils responded well to the new approach and would not have done so well with the traditional approach. That, of course, was countered by those pupils who they thought would have fared better with the traditional approach. Reassuringly things had not got worse, but there was no sign of significant improvement, things were much the same. So had the exercise been worthwhile or could it be considered as 'change for change's sake'?

They took the view that the change had been worthwhile despite the fact that the hoped for improvement had not materialised as they found other unanticipated benefits. They had enjoyed the change experience and working collectively on a curriculum development project. It had made them think about how they taught chemistry, developing new ideas and innovations to develop concepts and support pupils' learning. In short they felt they were better and more reflective chemistry teachers. This benefit can be considered as 'challenging the status quo'. Well rehearsed and effective practice delivered year on year can stifle innovation and make the change demanded by external forces (such as national curriculum developments) difficult to cope with – teachers become resistant to change. The best way to manage such change is to encourage a climate where teachers are supported to research and trial a manageable degree of change which they direct and control. Then they are more likely to be able to accommodate externally driven change.

Designing your own research

6

Education research should have a purpose; it should provide evidence for decision making that will lead to improvement. Research you do yourself need not be ground breaking or Earth shattering. Well founded incremental change can be more manageable and cumulatively can lead to significant change in practice and improvement.

The four approaches to educational research outlined above are not alternatives. Select and blend methods from all four as appropriate for what you want to study. Be critical and sceptical at the design stage of your study. Is your approach valid? Will your results be reliable? Have you chosen the most rigorous approach to collecting evidence that you can? In addition to your hands on research, do you need to read and research literature relevant to your study? Remember to apply the same critical evaluation and scepticism to the work of others that you apply to yourself. For example see Case Study Box 10 *Comparing SQA results with others*.

In your research seek the answers to questions that are relevant to your situation. Try new methods and materials in your classroom; set up trials, use control groups, make comparisons, use baseline studies.

Be critical of the questions you set out to answer; ask smart questions. Beware of broad open ended questions based on an observation; try to break down questions into hypotheses that have a prediction that can be tested. There can be several hypotheses for the same observation. For example, in a science department learners had been observed to struggle with the concept of density in floating and sinking experiments. The department decided to introduce measuring density to support learners' understanding. One group measured density as a mass to volume ratio while another group measured density as mass per unit volume. The impact of these two strategies on learners' understanding of density was then compared.

As well as being critical of the questions you set out to answer beware of making assumptions and generalising in your observations; examine your observations carefully. For example for decades concern has been raised on the gender bias in pupils' choice of science subjects. Typically the gender split in physics is 70/30 in favour of boys, in chemistry the split is generally 50/50 and in biology the gender split is 30/70 in favour of girls. Interestingly, until recently, the focus was often on the imbalance of girls studying physics; which in itself says something about the selective use of information! The outcome of this was often initiatives to recruit girls into physics which often resulted in more clever girls doing physics. What did not happen was a closer examination of the observed gender split. Boys and girls were treated as if they were two homogeneous groups. Is the gender split the same across the ability range? Do boys and girls who choose one, two or three science subjects show the same gender bias? If the bias in gender split across science subjects is to be addressed a closer examination of the observations on gender bias should lead to smarter questions being asked and useful hypotheses being tested.

CASE STUDY BOX 10

Comparing SQA results with others

Examining SQA results is changing from STACS (Standard Tables and Charts) to Insight, the new senior phase benchmarking tool [18]. STACS will continue to be available to schools as a source of historical data. Insight is available in state secondary schools; independent schools and colleges are not included. Insight makes comparisons to a 'virtual comparator' rather than between comparable schools. The virtual comparator is created by taking ten pupils with similar characteristics to each pupil in the school from the national population at random from outwith the local authority area. The characteristics are: gender, additional support needs (above or below 80% mainstream), leaving stage (S4 or S5/S6) and the Scottish Index of Multiple Deprivation (SIMD) based on post code.

At the present time (2015) school subject data in Insight allows comparisons to local authority and national data. A virtual comparator for subject comparisons is under development. For subject comparisons the virtual comparator will select a tenfold sample along the same lines as for the school comparison but from pupils who have studied the subject in question. In time relative values will be available for each subject. Relative value compares the performance of pupils in a particular course with the national performance in that course by pupils of similar general attainment. Pupils' general attainment is measured as the mean value of tariff points across all of the candidate's other courses within the same year and stage, excluding the course in question.

Making comparisons between subjects in a school is fraught with problems. The pupil cohorts taking subjects will be different and some subjects are more 'difficult' than others. Comparisons could be made on a matched pair design where groups of pupils who take the same two subjects (or more, although this is likely to reduce sample size) have their results compared. Another approach could be to compare performance between subjects with groups of similar attainment in all their other subjects using mean tariff scores. However the different difficulty level between subjects remains. SQA National Ratings provide a measure of the relative difficulty between subjects. National Ratings are not publically available but may be available by request from SQA (historical data on National Ratings can be found by entering SQA National Ratings in a search engine). Another measure of the relative difficulty of subjects is what was known in STACS as Progression Values. Progression Values give a measure of the correlation between performance in a subject at one SCQF level and the next SCQF level. For example an 'A' award at SCQF level 5 will show different levels of attainment at SCQF level 6 in different subjects.

The larger the number of pupils involved in the calculation the more reliance can be based on it. Hence why virtual comparators select ten equivalent pupils for each pupil in the sample school. Comparing two randomly selected virtual comparators should produce the same result. However if the school sample is relatively small, then the school values may deviate from the comparator as a consequence.

[18] Insight technical guidance, Insight help and support at http://insight-guides.scotxed.net/technical.htm (accessed May 2016).

6.1 ETHICS

Ethics is something that always has to be borne in mind when dealing with people. When gathering evidence on which to base educational improvement our approach to ethics should be sensible and proportionate. We should ask ourselves two basic questions: what is it we are asking learners to do?, and what are we going to do with the evidence we collect? The answers to these questions should guide our approach to ethical considerations.

Guidance on ethical educational research is provided by the British Educational Research Association [19]. It shares a set of principles comparable with other areas of research that involve people as the subjects or contexts of research. Although these guidelines are aimed at professional researchers carrying out commissioned research the principles also apply to teachers carrying out their own research. The cornerstones of ethical educational research are:

- voluntary informed consent
- right to withdraw
- best interests of the child
- confidentiality.

Where it is appropriate, voluntary informed consent should be obtained from pupils participating in educational research. Such consent will not always be necessary or appropriate. For example if a science department decided to extend the teaching of density to include measuring density to improve learners' understanding; they might measure density as a mass to volume ratio with one group of learners and as mass per unit volume with another group and compare the results. In such a case there would be no need to seek voluntary informed consent as the research and the evidence gathered can be considered as part of normal learning and teaching. However if a science department decided to investigate if senior pupils measuring their own blood pressure improved their understanding of blood pressure with an experimental and a negative control group; then students allocated to the experimental group should give informed consent voluntarily. In this case verbal consent from the students would be sufficient in line with current SSERC advice and practice on pupils as subjects of experiment or investigation [20]. In general where the research is based on observations of classroom behaviour that teachers would normally see as part of learning and teaching, there is no need to seek voluntary informed consent. Observing classroom behaviour in this way is also in line with the British Psychological Society's Code of Ethics and Conduct [21].

There may be occasions where obtaining voluntary informed consent (if it is considered necessary) is excluded by the nature of the research. For example if there was reason to suspect that there may be a placebo or Hawthorne effect or if blind or double blind trials were being conducted. In research such conditions are usually examined by an ethics committee. In this regard, interesting parallels can be drawn with medical research; see for example, Case Study Box 11 *Ethics in Medical Trials*. In most cases of school based research it would be sensible and sufficient to inform and discuss such a situation with peers and with the Head Teacher who is best placed to seek further advice if necessary. SSERC is happy to assist with such advice if required.

[19] Ethical Guidelines for Educational Research, British Educational Research Association (BERA), London, 2011 at https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf (accessed May 2016).

[20] Materials of Living Origin – Educational Uses – A Code of Practice for Scottish Schools and Colleges, SSERC, 2012 at http://www.sserc.org.uk/images/Publications/Biology/SSERC-Materials_of_Living_Origin_Code_of_Practice.pdf (accessed May 2016).

[21] Code of Ethics and Conduct, British Psychological Society, Leicester, 2009 at http://www.bps.org.uk/system/files/documents/code_of_ethics_and_conduct.pdf (accessed May 2016).

CASE STUDY BOX 11

Ethics in medical trials

When considering randomised controlled trials for approval, medical ethical committees often have to weigh up the consequences of having one group which receives a placebo in a double blind randomised controlled trial. Some might consider the inclusion of a placebo group unethical, while others would argue that if the outcome of the treatment being trialled is unknown it is unethical not to conduct a trial. This dilemma is well illustrated by the CRASH trial [22]. For many years it was common to treat patients with a head injury with a steroid injection. This was based on the knowledge that steroids reduce swelling and as a head injury causes the brain to swell it made sense to give steroids. Some doctors gave steroids based on this belief and others did not. The CRASH trial was designed to determine the effect of steroid injection on brain injury. Ethics committees had difficulty in both approving a placebo treatment and randomising unconscious patients to a treatment or control group. When the trial was eventually approved and conducted it turned out that the patients receiving the steroid injection fared worse than the control group. More deaths were recorded (2.5 per hundred) in the treatment group.

A current trial (due to report in 2019) is assessing the use of adrenaline in patients with out of hospital cardiac arrest [23]. Adrenaline is known to increase blood flow to the heart helping to restart the heart. However, the effect of adrenaline in the longer term for such patients may cause harmful side effects. In the trial patients will randomly receive adrenaline or a saline placebo in a double blind trial. Patients surviving will be informed of their enrolment in the trial in hospital, at which point their consent to continue in the trial will be sought. This trial has been given ethical approval.

The right to withdraw should be respected. In most cases of school based educational research, seeking (and approving) the right to withdraw would only take place in exceptional circumstances as the research will be conducted as part of everyday learning and teaching. The right to withdraw can also apply to learning activities more generally as a result of culture and/or beliefs. Such instances should be dealt with the respect and sensitivity accorded to those beliefs and culture.

Research should be conducted in the best interests of the child. That is the purpose of the research should be of benefit to those participating and to their future peers. The research should be part of a culture of educational improvement. Results should be treated with confidentiality and not attributed to individuals. Participants should be entitled to know their own results and/or how the data is used. Teachers should be open to explaining and sharing the results of their research with their peers, pupil participants and parents/carers.

[22] Bad Pharma, Ben Goldacre, Fourth Estate, London, 2013, pp 232-234.
[23] Randomised placebo controlled trial of adrenaline for out of hospital cardiac arrest (Paramedic 2) HTA - 12/127/126 at http://www.nets.nihr.ac.uk/projects/hta/12127126 (accessed May 2016).

7 In conclusion

A survey of MPs' attitudes by Ipsos MORI [24] found support for using randomised controlled trials (RCTs) to test social policy. It showed that MPs:

- Support the use of controlled trials to design and test social policy
- Expect that the use of trials in policy will increase
- Don't consider the cost of running policy trials a barrier.

However it also found MPs had concerns over the fairness of allocating people to groups at random, and showed some confusion about the purpose of control groups. It also showed that MPs are more likely at present to pay attention to and base their decisions on:

- Evidence from experts (e.g. academics or think tanks)
- Views of constituents
- Views of practitioners (e.g. doctors, teachers, police etc.)
- Personal experience/principles
- Pilot schemes without control groups.

This survey encapsulates the work that has to be done before evidence based decision making can be embedded in educational practice. This is no different to what happened in agriculture in the 1920s and 1930s and in medicine in the 1970s and 1980s. Teachers can help to effect a change to evidence based decision making by carrying out small scale randomised controlled trials, well conducted correlation studies and social research and presenting their results to support argument and debate when engaging with those in authority.

Image: Sandol/www.morguefile.com

[24] What do MPs think of randomised controlled trials? Sense about science, 2014 at http://www.senseaboutscience.org/pages/what-do-mps-think-of-rcts.html (accessed May 2016).

Further reading

8

If you wish to learn more about evidence based education you should find the following publications useful. The first three papers in this list are relatively short and do not take long to read. The first two deal with randomised controlled trials and the third with qualitative research methods.

- *Building Evidence into Education*, Ben Goldacre (2013) at http://www.tactyc.org.uk/pdfs/Goldacre-Paper.pdf (accessed May 2016).
  This paper makes the case for the use of evidence based practice in education to improve outcomes for learners and to empower teachers to make decisions about what works best. It also discusses the structures that need to be in place to ensure that the results of good quality research are disseminated and put into practice.

- *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*, Laura Haynes, Owain Service, Ben Goldacre, David Torgerson, (2012), Cabinet Office Behavioural Insights Team https://www.gov.uk/government/uploads/system/uploads/attachment _ data/file/62529/TLA-1906126.pdf (accessed May 2016).
  This paper describes how to design and carry out a randomised controlled trial to test if a policy is working. It also discusses the analysis of results and their use to improve policy.

- *Reflections on a personal journey in research*, Wynne Harlen, ASE, 2009 at http://www.ase.org.uk/resources/scitutors/research/research-wynne-harlen/ (accessed May 2016).
  This paper makes the case for qualitative research in providing evaluation of learning strategies and interventions.

The following books deal with the importance of using evidence from randomised controlled trials when making medical and health care decisions. The case they make for basing decisions on sound research evidence is also applicable to education and other areas of public service making these books of relevance to those interested in educational improvement.

- *Testing Treatments – Better Research for Better Health Care*, Imogen Evans, Hazel Thornton & Iain Chalmers, Pinter and Martin Ltd, London, 2010 at http://www.testingtreatments.org/ (accessed May 2016).
  Available as a free pdf download, this book not only makes the case for evidence based medicine but for the results of research being made available to patients and the wider public in an honest and accessible form.

- *Bad Science*, Ben Goldacre, Fourth Estate, London, 2008.
  The purpose behind this book is to educate the general public so that they can become critical consumers of scientific research – teaching good science by examining the bad.

- *Bad Pharma*, Ben Goldacre, Fourth Estate, London, 2013.
  This book uses evidence based science to attack the abuse of evidence by the pharmaceutical industry to mislead about the effectiveness of medical treatments.

SSERC (Scottish Schools Education Research Centre), 2 Pitreavie Court, South Pitreavie Business Park, Dunfermline KY11 8UU.
Telephone 01383 626 070, fax 01383 842 793, e-mail sts@sserc.org.uk, website www.sserc.org.uk

SSERC is a Company Limited by Guarantee (Scottish Company No. SC131509) and a registered educational charity
(Scottish Charity No. SCO17884) Registered Office – 5th Floor, Quartermile Two, 2 Lister Square, Edinburgh EH3 9GL.

ISBN 978-0-9933282-0-6.