National College for Teaching & Leadership

# Closing the Gap: Test and Learn

## Research Development and Networking Event 4

## Option A – a one day event explaining how to design, implement, analyse and write up a piece of teacher led-experimental research

In partnership with

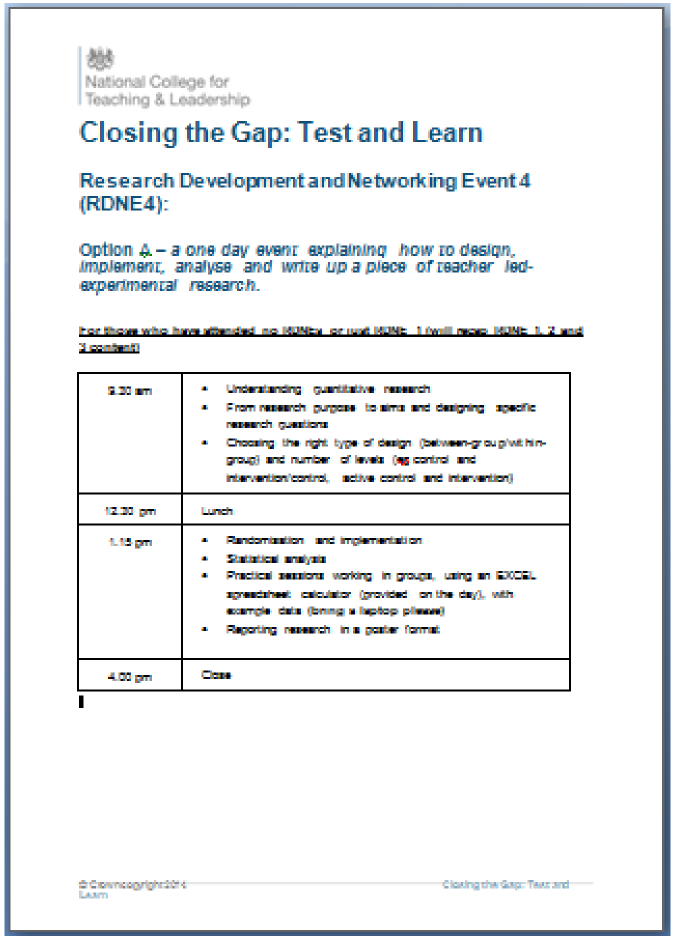
CfBT Education Trust

curee

University of Oxford

National College for Teaching & Leadership

- This event is for those who have attended no RDNEs or just RDNE 1 (will recap RDNE 1, 2 and 3 content)


National College for Teaching & Leadership

## Closing the Gap: Test and Learn

**Research Development and Networking Event 4 (RDNE4):**

**Option A – *a one day event explaining how to design, implement, analyse and write up a piece of teacher led-experimental research.***

**For those who have attended no RDNEs or just RDNE 1 (will recap RDNE 1, 2 and 3 content)**

| | |
|---|---|
| 9.30 am | • Understanding quantitative research<br>• From research purpose to aims and designing specific research questions<br>• Choosing the right type of design (between-group/within-group) and number of levels (eg control and intervention/control, active control and intervention) |
| 12.30 pm | Lunch |
| 1.15 pm | • Randomisation and implementation<br>• Statistical analysis<br>• Practical sessions working in groups, using an EXCEL spreadsheet calculator (provided on the day), with example data (bring a laptop please)<br>• Reporting research in a poster format |
| 4.00 pm | Close |

National College for Teaching & Leadership

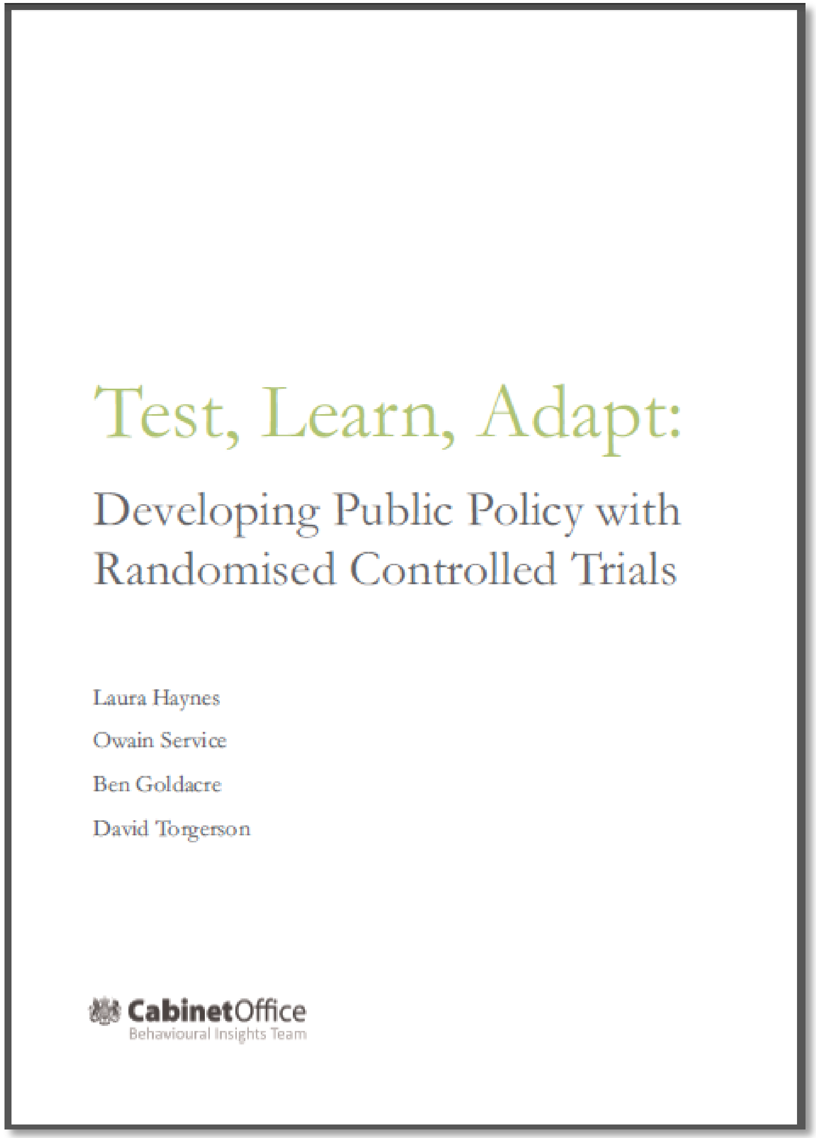
# Meet each other again or for the first time

- On your tables, introduce yourself to someone you do not know

- What do you want to get out of today?

National College for
Teaching & Leadership

# Why randomised controlled trials?


Test, Learn, Adapt:

Developing Public Policy with
Randomised Controlled Trials

Laura Haynes
Owain Service
Ben Goldacre
David Torgerson

CabinetOffice
Behavioural Insights Team

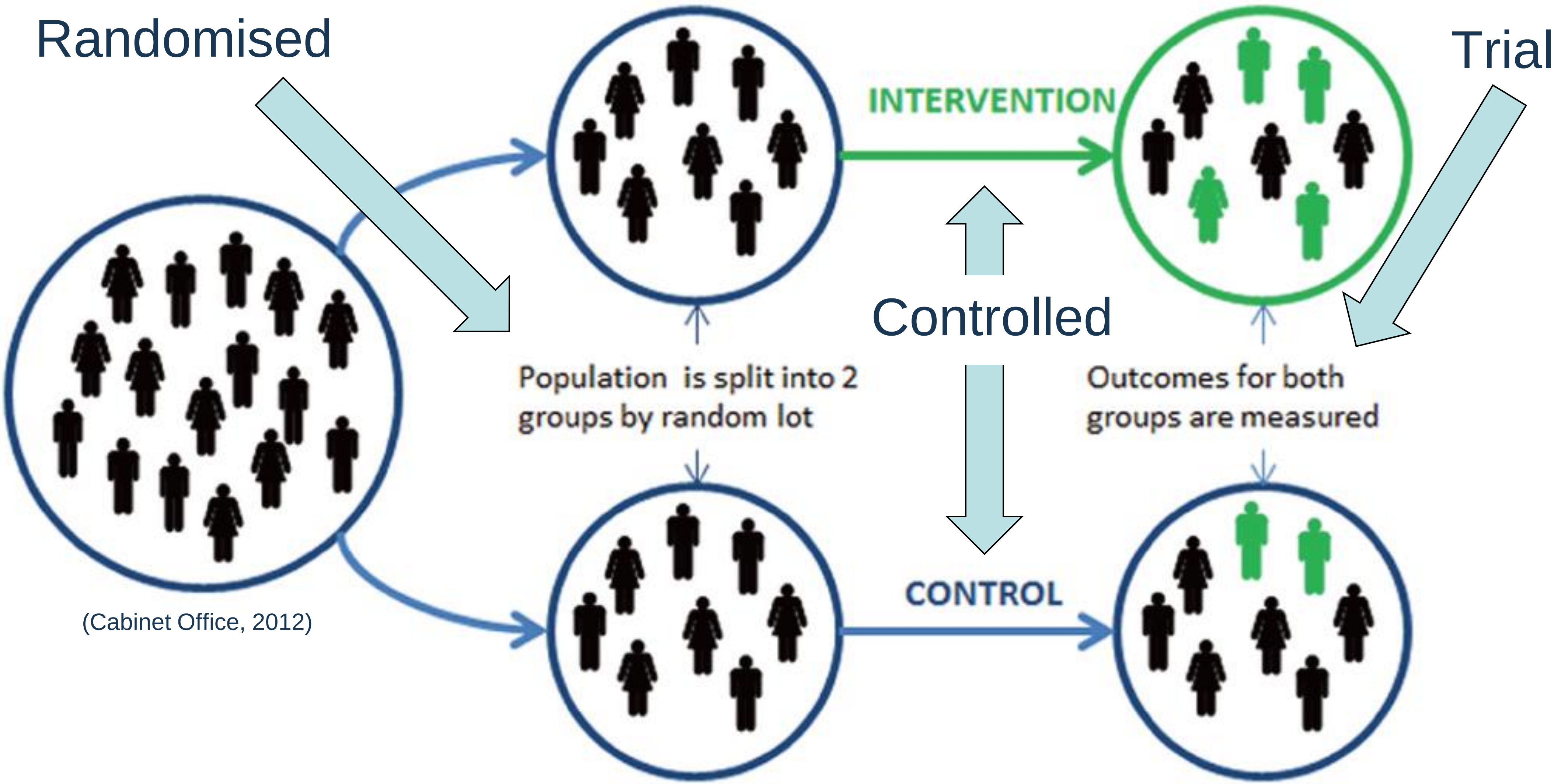
- Introduction of a randomly assigned **control group** enables the effectiveness of a new **intervention** to be compared to what would have happened if you had changed nothing

- Eliminates a whole host of biases that normally complicate evaluation

(Cabinet Office, 2012)

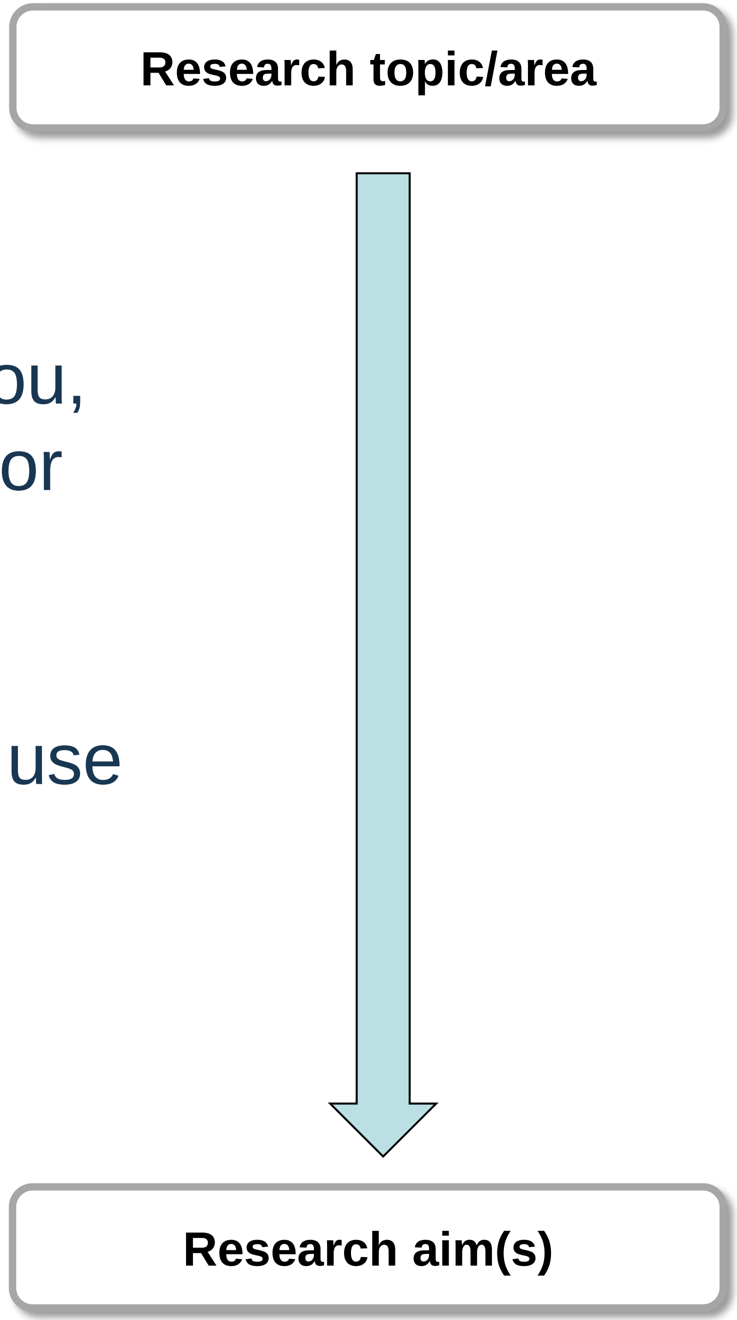
National College for
Teaching & Leadership

# Reminding ourselves of the basic approach




(Cabinet Office, 2012)

National College for
Teaching & Leadership



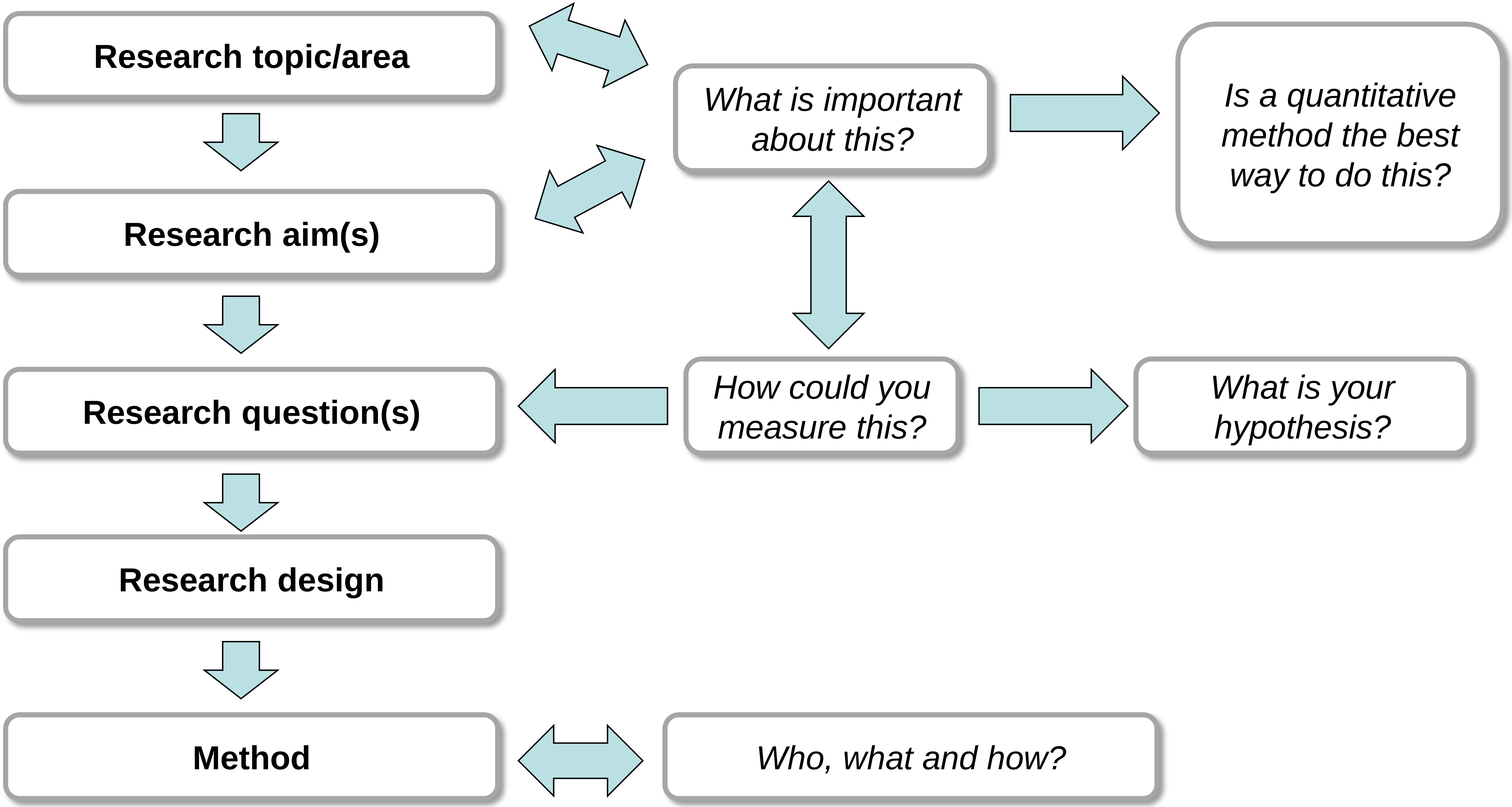

## Step 1

- Identify something that really interests you, that you think would make a good area for research using an RCT
- What is important about that?
- Is a quantitative method the right one to use for this research, or would a qualitative method be better? (Can differences be measured in some way?)

## Step 2

- What is your research aim(s)?

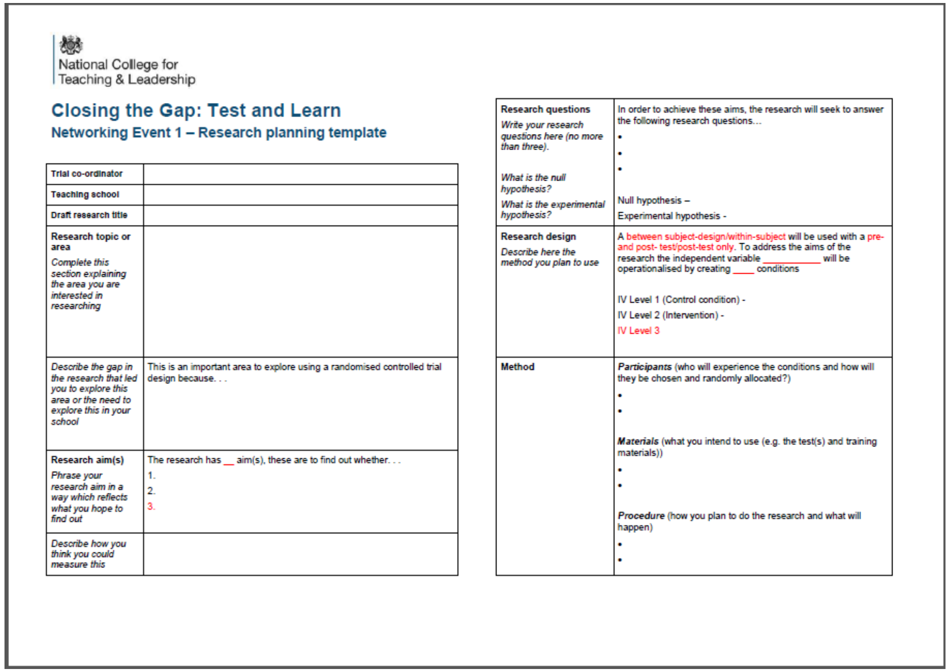
# The planning stages



National College for Teaching & Leadership

# In groups

- Think of a research topic or area that you would really like to have an answer to
- Move to work with people with similar interests
- There is a design template to help you. You will use this as a group this morning and design an RCT



National College for Teaching & Leadership

## Closing the Gap: Test and Learn

### Networking Event 1 – Research planning template

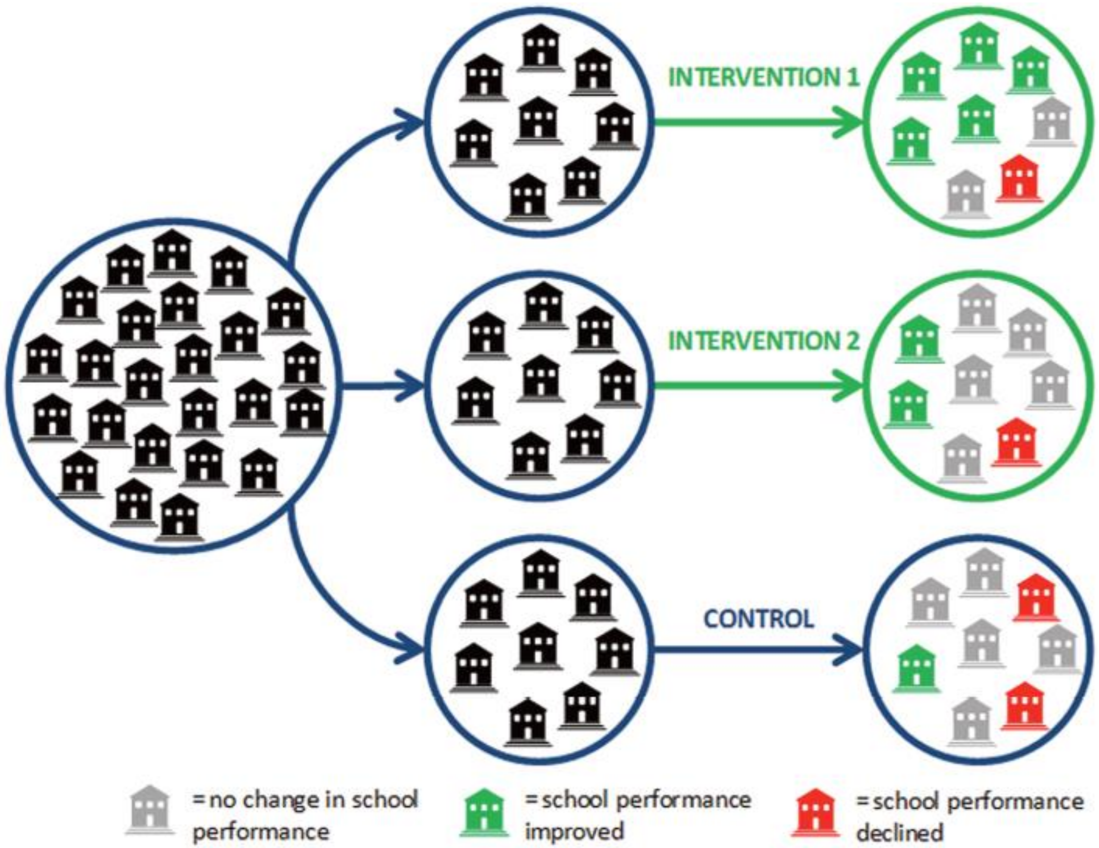
| | |
|---|---|
| **Trial co-ordinator** | |
| **Teaching school** | |
| **Draft research title** | |
| **Research topic or area**<br>*Complete this section explaining the area you are interested in researching* | |
| *Describe the gap in the research that led you to explore this area or the need to explore this in your school* | This is an important area to explore using a randomised controlled trial design because. . . |
| **Research aim(s)**<br>*Phrase your research aim in a way which reflects what you hope to find out* | The research has __ aim(s), these are to find out whether. . .<br>1.<br>2.<br>3. |
| *Describe how you think you could measure this* | |

| | |
|---|---|
| **Research questions**<br>*Write your research questions here (no more than three).*<br>*What is the null hypothesis?*<br>*What is the experimental hypothesis?* | In order to achieve these aims, the research will seek to answer the following research questions…<br>•<br>•<br>•<br>Null hypothesis –<br>Experimental hypothesis - |
| **Research design**<br>*Describe here the method you plan to use* | A between subject-design/within-subject will be used with a pre- and post- test/post-test only. To address the aims of the research the independent variable ___________ will be operationalised by creating ____ conditions<br>IV Level 1 (Control condition) -<br>IV Level 2 (Intervention) -<br>IV Level 3 |
| **Method** | ***Participants*** (who will experience the conditions and how will they be chosen and randomly allocated?)<br>•<br>•<br>***Materials*** (what you intend to use (e.g. the test(s) and training materials))<br>•<br>•<br>***Procedure*** (how you plan to do the research and what will happen)<br>•<br>• |

National College for Teaching & Leadership

# Common variations on the basic RCT structure

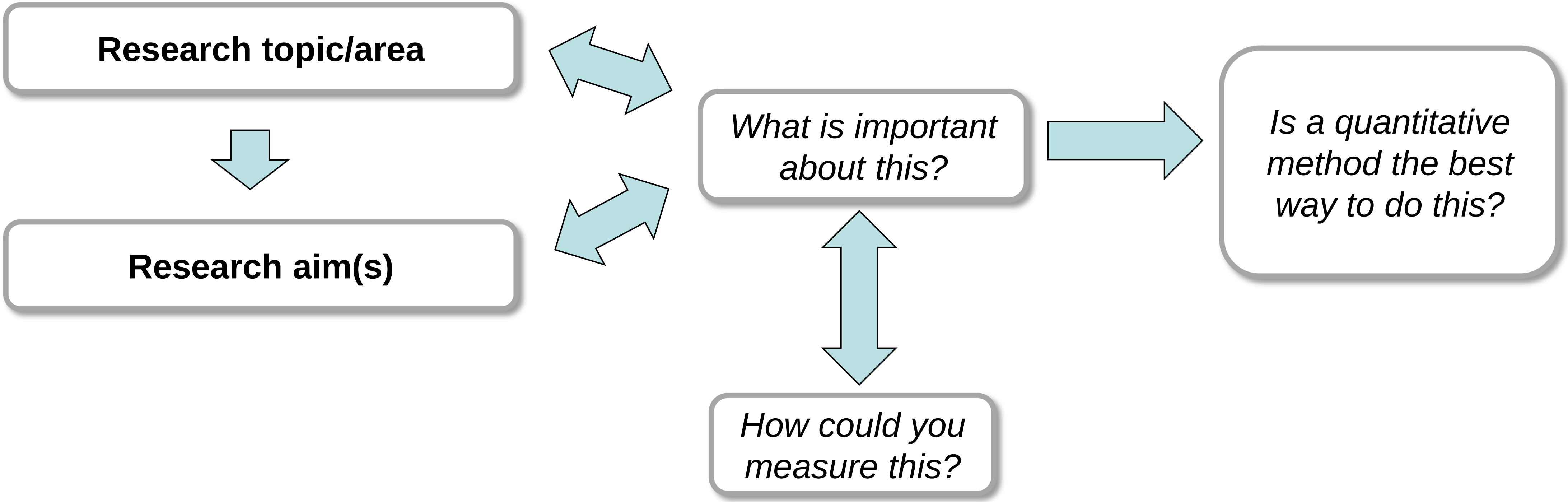
- Sometimes an RCT might test two **interventions** at once and compare them to a **control group**
- This could be a second **intervention** or an ‘active control’
- Any form of RCT could also include a **before test** as well as an **after test**




(Cabinet Office, 2012)

National College for Teaching & Leadership

# Now look at the research aim(s) section and decide on a way of measuring this – discuss



# Types of research design

- **Experiment** – random allocation to control or intervention
- **Quasi-experiment** – a comparison of different existing groups (eg in research wanting to know if boys do better in group work than girls; here you would randomly sample boys and girls from the population)

↓

- **Between-subject design** – different participants experience different things (eg they are in the control **or** intervention)
- **Within-subject design** – participants experience the same things but in a different order (eg control→intervention **or** intervention→control)

# Key terminology: dependent and independent variables

**Independent variable (IV)** – what you manipulate as the experimenter (for example, you might have two ‘levels’ to your IV – a control condition and an intervention)

**Dependent variable(s) (DV)** – The results which ‘depend’ on your IV manipulation. In other words, the measurement(s) you use (practically, you need to keep the measure(s) constant for each level of your IV)

**Research question(s)** ← *How could you measure this?* → *What is your hypothesis?*

# Hypotheses

**Null hypothesis:** Three lessons of group work does not improve maths attainment as measured by the number of column addition problems solved correctly in a 10-minute test

**Experimental hypothesis:** Three lessons of group work improves maths attainment as measured by the number of column addition problems solved correctly in a 10-minute test

**Research question(s)** ← *How could you measure this?* → *What is your hypothesis?*

# Research question

- Does group work improve column addition problem-solving ability?

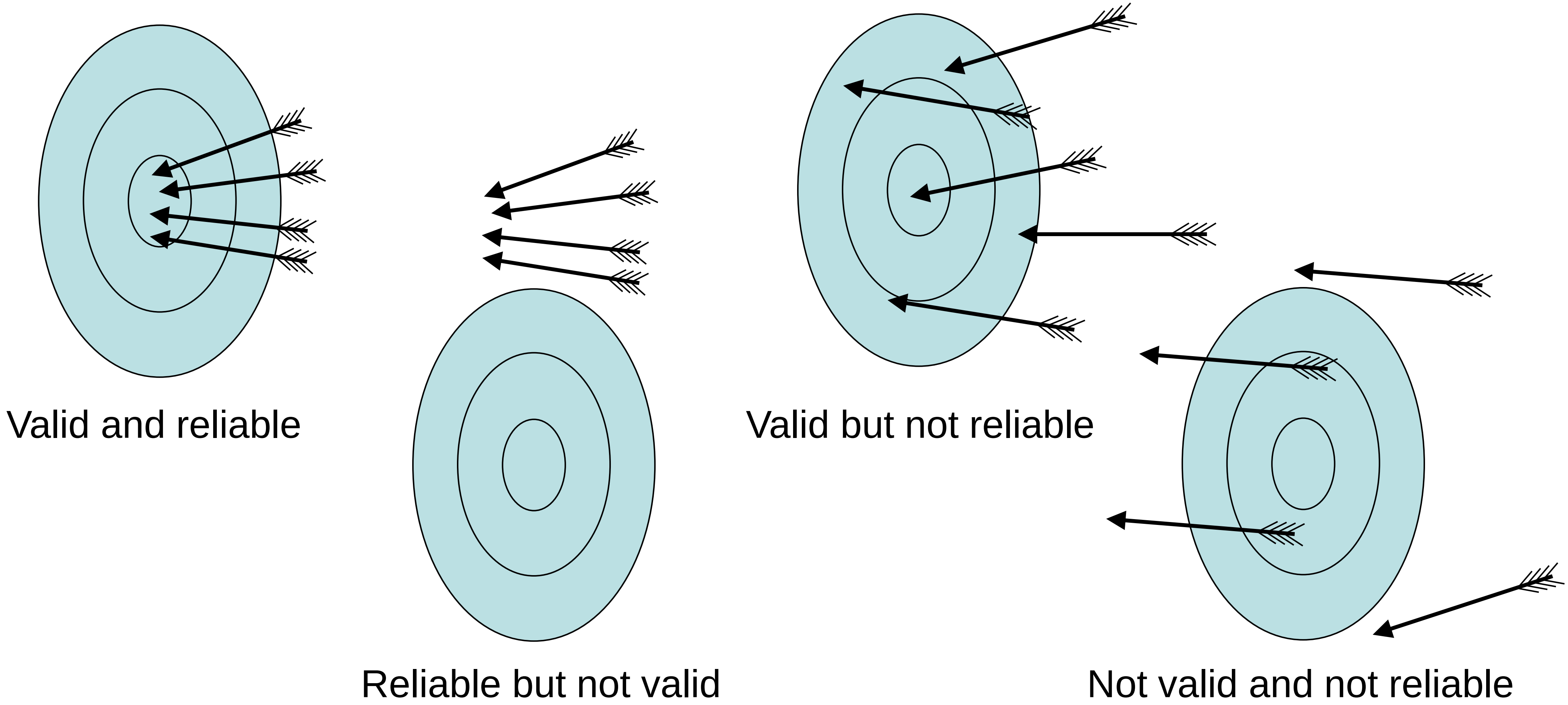
National College for Teaching & Leadership

# Deciding on the dependent variable(s)

- Look at your **research question(s)**
- What measure(s) could you use to answer the question(s)?
- You could apply an existing test you already use, or design a new one (called '**primary data**') if you collect the data directly in both cases
- Or you could use existing school improvement data and information that is already being collected (**secondary data**)
- Or a mixture of **primary and secondary data**

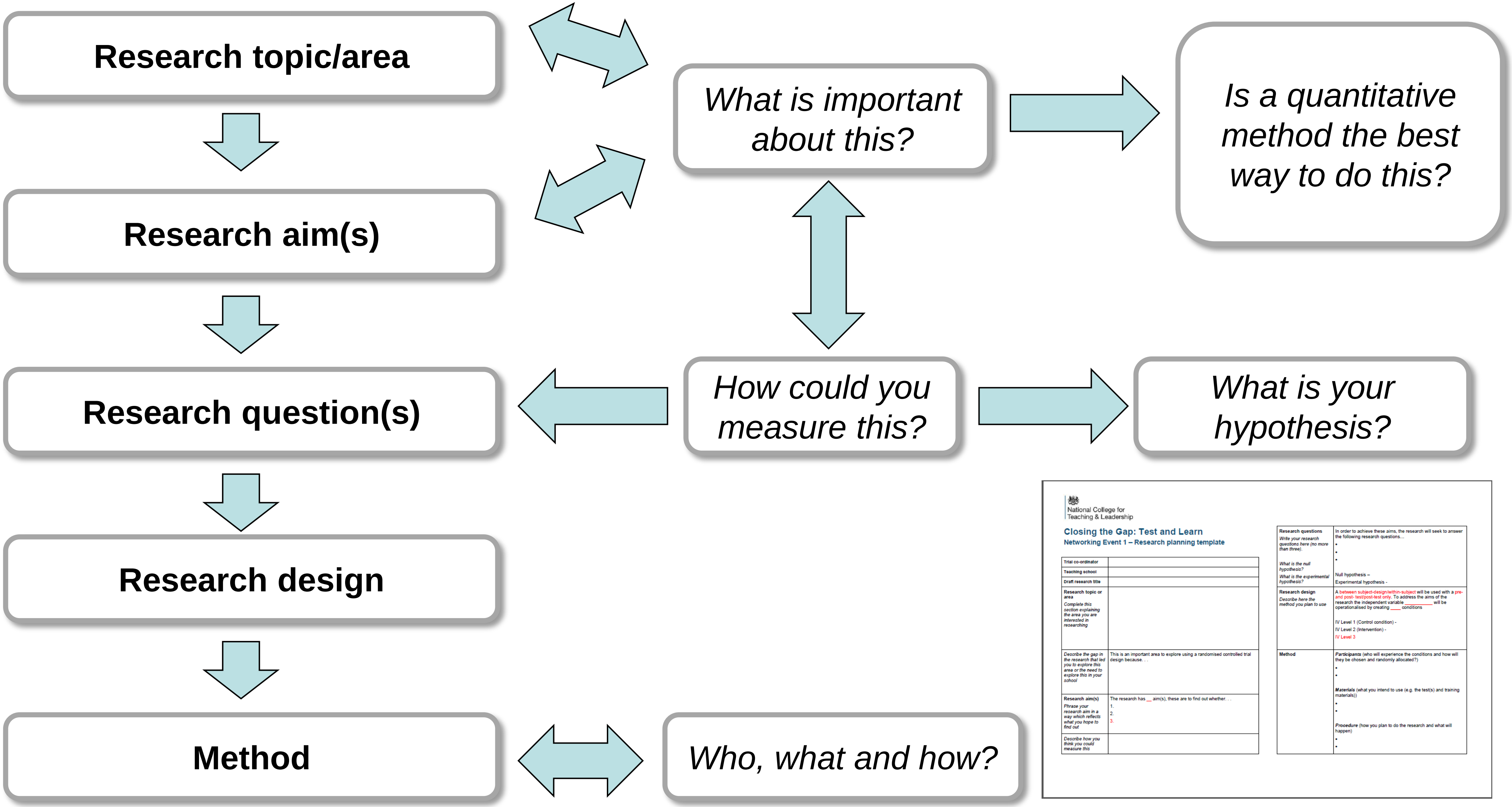
# Validity and reliability

- Is your measure **valid** and **reliable**? Imagine the centre of the target is your research aim – does you DV relate to your IV?



National College for Teaching & Leadership

# Back to the template




National College for Teaching & Leadership

**Closing the Gap: Test and Learn**

**Networking Event 1 – Research planning template**

| | |
|---|---|
| **Trial co-ordinator** | |
| **Teaching school** | |
| **Draft research title** | |
| **Research topic or area** *Complete this section explaining the area you are interested in researching* | |
| *Describe the gap in the research that led you to explore this area or the need to explore this in your school* | This is an important area to explore using a randomised controlled trial design because. . . |
| **Research aim(s)** *Phrase your research aim in a way which reflects what you hope to find out* | The research has __ aim(s), these are to find out whether. . .<br>1.<br>2.<br>3. |
| *Describe how you think you could measure this* | |

| | |
|---|---|
| **Research questions** *Write your research questions here (no more than three).*<br>*What is the null hypothesis?*<br>*What is the experimental hypothesis?* | In order to achieve these aims, the research will seek to answer the following research questions...<br>•<br>•<br>•<br>Null hypothesis –<br>Experimental hypothesis - |
| **Research design** *Describe here the method you plan to use* | A between subject-design/within-subject will be used with a pre- and post- test/post-test only. To address the aims of the research the independent variable __________ will be operationalised by creating ____ conditions<br>IV Level 1 (Control condition) -<br>IV Level 2 (Intervention) -<br>IV Level 3 |
| **Method** | ***Participants*** (who will experience the conditions and how will they be chosen and randomly allocated?)<br>•<br>•<br>***Materials*** (what you intend to use (e.g. the test(s) and training materials))<br>•<br>•<br>***Procedure*** (how you plan to do the research and what will happen)<br>•<br>• |

# Between-subject designs (or independent measures)

## Main advantages

- Use when the effects of your intervention are irreversible
- Reduces chance of participants becoming bored
- Removes risk of becoming better simply through practice

## Main disadvantages

- Needs larger number of participants
- Variability between participants can affect results

National College for Teaching & Leadership

# Within-subject designs (or repeated measures)

## Main advantages

- Requires fewer participants (typically half the number)
- Reduces error associated with individual differences

## Main disadvantages

- Fatigue
- Order effects – the effect on the second condition of having done the first one (eg performance improvement), usually dealt with by **counterbalancing**

# When to test

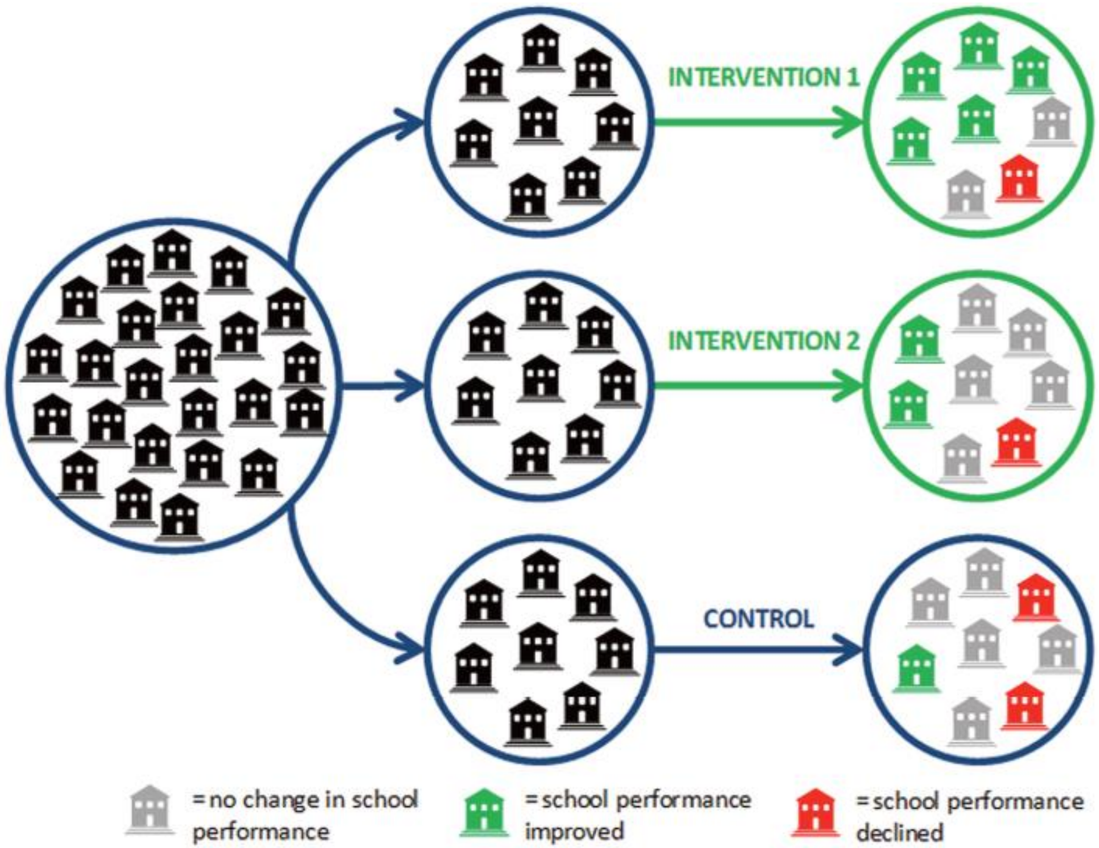
The main options:

- **Post-test (after test)** – only test the participants at the end of the procedure

- **Pre-test and post-test (before and after test)** – test the participants before and after the procedure

# Do you need a third condition?

- You might decide that you need more than one **condition** (if you have three **conditions** we would say that your **independent variable** has **three levels**)
- The other **condition** could be an ‘active control’ or placebo, or something else you want to test at the same time




(Cabinet Office, 2012)

National College for Teaching & Leadership

# Hypotheses

**Null hypothesis:** Three lessons of group work does not improve maths attainment as measured by the number of column addition problems solved correctly in a 10-minute test

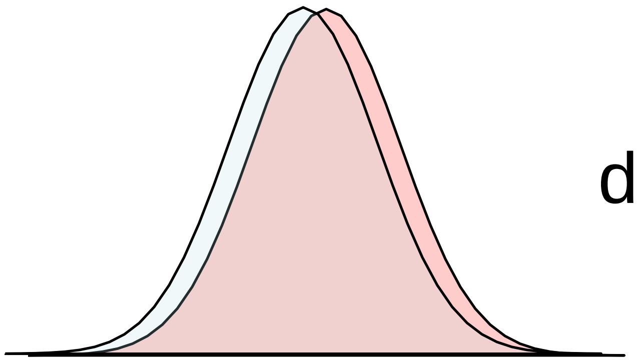
**Experimental hypothesis:** Three lessons of group work improves maths attainment as measured by the number of column addition problems solved correctly in a 10-minute test

These represent statistical thresholds expressed by the level of significance obtained from the final analysis (for example, $p < .05$, if a one-tailed hypothesis (more about this later))

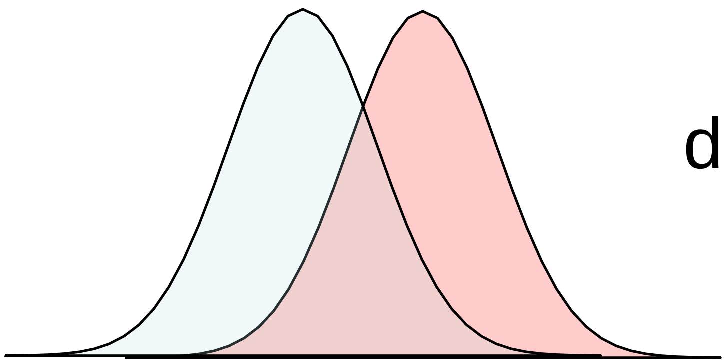
# Line up activity to help understand what is going on with Cohen's d

National College for Teaching & Leadership

# For example



d = 0.1



d = 0.8

| Effect Size d | Percentage of pupils likely to have been clearly affected by the intervention, if the result is significant (percentage of non-overlap between distributions) |
|---|---|
| 1.3 | 65.3% |
| 1.2 | 62.2% |
| 1.1 | 58.9% |
| 1.0 | 55.4% |
| 0.9 | 51.6% |
| 0.8 | 47.4% |
| 0.7 | 43.0% |
| 0.6 | 38.2% |
| 0.5 | 33.0% |
| 0.4 | 27.4% |
| 0.3 | 21.3% |
| 0.2 | 14.7% |
| 0.1 | 7.7% |
| 0.0 | 0% |

# Some statistics

- Effect size – difference in mean, taking into account the spread of scores (or variance). One way to do this is with Cohen's *d*.

$$Effect\ size\ d = \frac{Intervention\ mean - Control\ mean}{pooled\ Standard\ Deviation}$$

- Cohen's *d* has limitations which we will cover this afternoon (if your data is not normally distributed, for example)

- *p* level – probability that the change happened by chance (for example, we might look for a minimum of $p < .05$ in order to reject the null hypothesis); you can calculate in Excel.

National College for Teaching & Leadership

# *p* (significance) is a function of effect size and sample size

- If you do not have a large enough sample you risk ending up with an effect that is not significant (i.e. does not reach a level of significance $p < .05$) when there really is one

- To detect a $d = 0.4$ effect size (between groups) you need 78 in each group if you think you know which way the mean will go (**one-tailed hypothesis**) and 100 in each group if you do not (**two-tailed hypothesis**)

- Working out sample size first is called **power analysis** (prior research may suggest the effect size to use, you could use Hattie's tables (Hattie, 2009))

Hattie, J. (2009) *Visible learning: a synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

# But keeping it simple

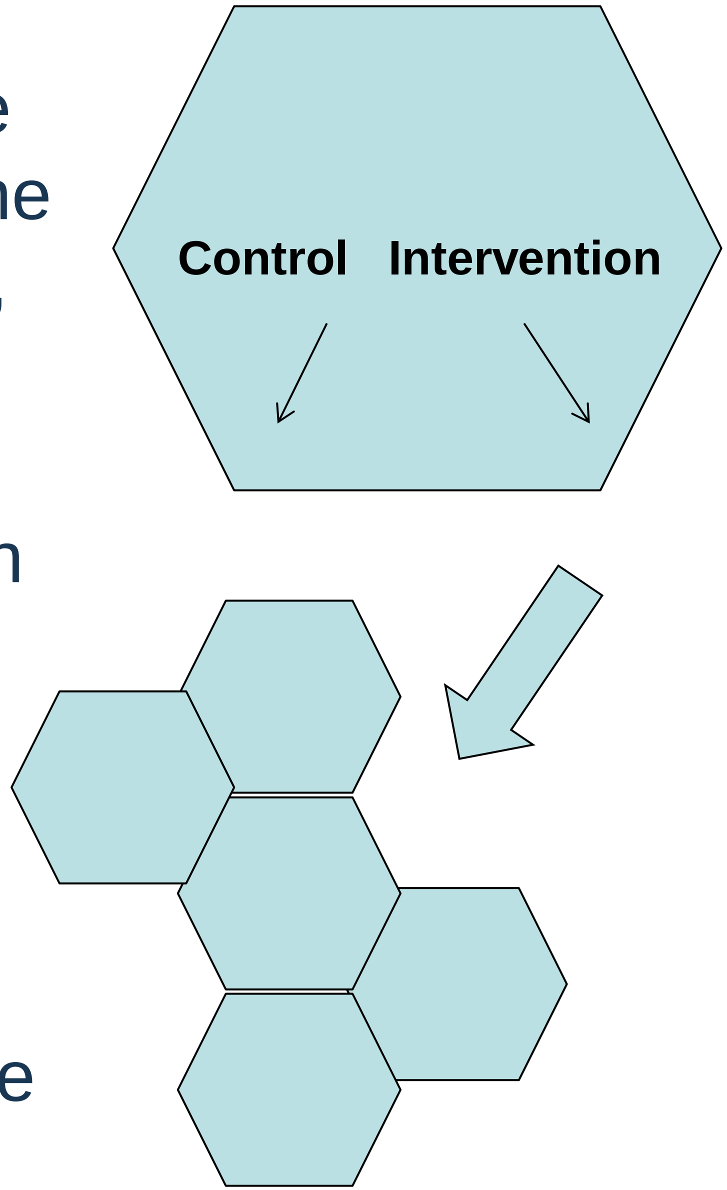
- For the purposes of your current activity, if you were going to look at a final post-test between-subject effect you would need:

A. 156 children (78 in each group) to detect a *d* = 0.4 effect size with a 5 in 100 probability that the results were not arrived at by chance, where you think the mean will go one way.

B. 200 children (100 in each group) to detect a *d* = 0.4 effect size with a 5 in 100 probability that the results were not arrived at by chance, where you do not know which way the mean will go.

# Building your sample size

- Some designs could be carried out where a large group of children are blocked in the timetable (eg a trial in a single PE lesson, or across a series of blocked maths lessons)
- Alternatively, you could take the approach often used in cognitive psychology and 'think like a bee', by keeping the protocol consistent and amalgamating lots of consistent cells of delivery (eg in a single lesson design where children were randomly allocated to different tasks in the same lesson; or two parallel classes)

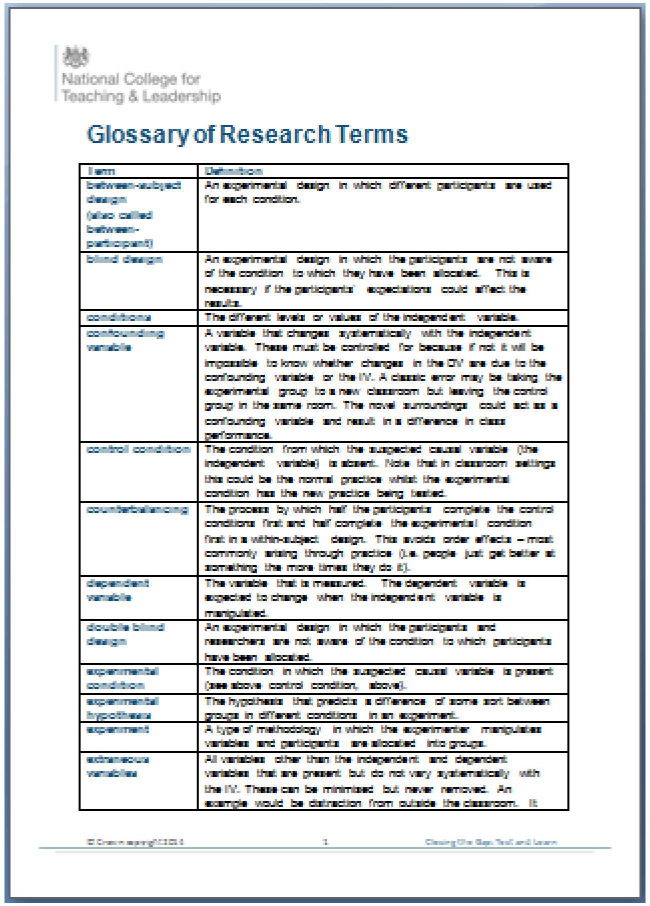


National College for
Teaching & Leadership

# Looking out for things that might affect your results

**Extraneous variable** – does not vary systematically with the IV. Can be minimised but never removed (for example, distraction from outside the classroom may be present in both your conditions and therefore not affect your experiment results but you can minimise the impact on the experiment (in this case, by shutting windows)).

**Confounding variable** – a variable that changes systematically with the independent variable. These must be controlled for because if not it will be impossible to know whether changes in the DV are due to the confounding variable or the IV.


National College for
Teaching & Leadership

Glossary of Research Terms

| Term | Definition |
|---|---|
| between-subject design (also called between-participant) | An experimental design in which different participants are used for each condition. |
| blind design | An experimental design in which the participants are not aware of the condition to which they have been allocated. This is necessary if the participants' expectations could affect the results. |
| conditions | The different levels or values of the independent variable. |
| confounding variable | A variable that changes systematically with the independent variable. These must be controlled for because if not it will be impossible to know whether changes in the DV are due to the confounding variable or the IV. A classic error may be taking the experimental group to a new classroom but leaving the control group in the same room. The novel surroundings could act as a confounding variable and result in a difference in class performance. |
| control condition | The condition from which the suspected causal variable (the independent variable) is absent. Note that in classroom settings this could be the normal practice whilst the experimental condition has the new practice being tested. |
| counterbalancing | The process by which half the participants complete the control conditions first and half complete the experimental condition first in a within-subject design. This avoids order effects – most commonly arising through practice (i.e. people just get better at something the more times they do it). |
| dependent variable | The variable that is measured. The dependent variable is expected to change when the independent variable is manipulated. |
| double blind design | An experimental design in which the participants and researchers are not aware of the condition to which participants have been allocated. |
| experimental condition | The condition in which the suspected causal variable is present (see above control condition, above). |
| experimental hypothesis | The hypothesis that predicts a difference of some sort between groups in different conditions in an experiment. |
| experiment | A type of methodology in which the experimenter manipulates variables and participants are allocated into groups. |
| extraneous variables | All variables other than the independent and dependent variables that are present but do not vary systematically with the IV. These can be minimised but never removed. An example would be distraction from outside the classroom. It |

National College for Teaching & Leadership

# Group discussion

- Consider a design you are thinking of using, or one of the ones we have talked about

- How could you control some of the extraneous variables (often these are environmental factors) in the procedure you use?

- Or collect a statistical measure to check to see if there was an effect when you do your analysis?

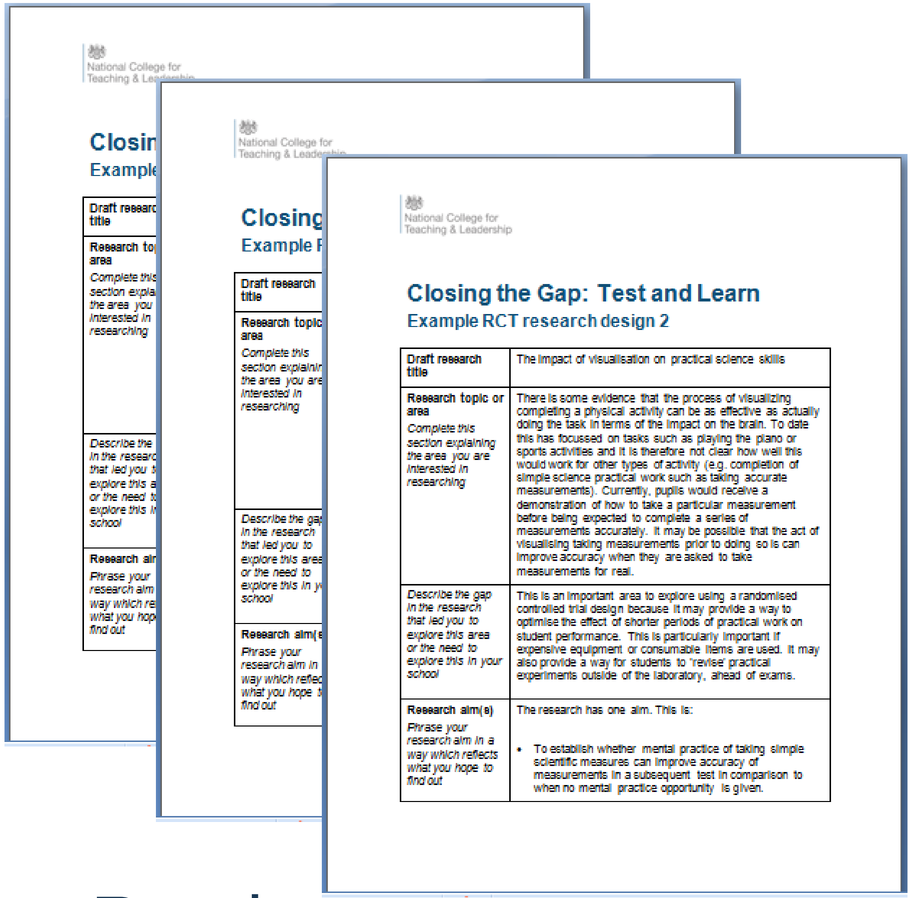
**Think about:**

- Teacher variation
- School day
- Response change because children know they are in a trial
- Way the intervention is applied and used
- Nature of the control group activities

…and anything else you can think of...

# Critical analysis

- If you have brought a design you can use this, or:
- In your packs you have three RCT examples
- In groups, choose one
- Analyse in relation to limitations/issues in the design and procedures described and suggest some alternatives



National College for Teaching & Leadership

Closing the Gap: Test and Learn

Example RCT research design 2

| | |
|---|---|
| **Draft research title** | The impact of visualisation on practical science skills |
| **Research topic or area** *Complete this section explaining the area you are interested in researching* | There is some evidence that the process of visualizing completing a physical activity can be as effective as actually doing the task in terms of the impact on the brain. To date this has focussed on tasks such as playing the piano or sports activities and it is therefore not clear how well this would work for other types of activity (e.g. completion of simple science practical work such as taking accurate measurements). Currently, pupils would receive a demonstration of how to take a particular measurement before being expected to complete a series of measurements accurately. It may be possible that the act of visualising taking measurements prior to doing so is can improve accuracy when they are asked to take measurements for real. |
| *Describe the gap in the research that led you to explore this area or the need to explore this in your school* | This is an important area to explore using a randomised controlled trial design because it may provide a way to optimise the effect of shorter periods of practical work on student performance. This is particularly important if expensive equipment or consumable items are used. It may also provide a way for students to 'revise' practical experiments outside of the laboratory, ahead of exams. |
| **Research aim(s)** *Phrase your research aim in a way which reflects what you hope to find out* | The research has one aim. This is: • To establish whether mental practice of taking simple scientific measures can improve accuracy of measurements in a subsequent test in comparison to when no mental practice opportunity is given. |

- Read
- Discuss
- Present back in groups

# Randomisation and EXCEL randomisation activity

National College for Teaching & Leadership

# Randomisation is the distinguishing feature of an RCT

- **Random sampling** is different to **random allocation**
- **Random sampling** improves the external and 'ecological' validity of a study – random sampling means that a participant group is more likely to represent the population
- **Random allocation** – does not improve **external validity** but does improve **internal validity** by helping to ensure that the results are accurate for the group that was chosen
- **Random sampling** first, then **random allocation**, is the strongest form of practice as it removes biases both in the sampling process and in allocation to levels of the IV (or the order in which the levels of the IV will be experienced in a within-subject design)

National College for Teaching & Leadership

# Types of randomisation and their limitations

**Simple randomisation** (could be done with a coin toss/rolling a dice, or using the RAND function in Excel)

- In Excel, generate random numbers (in the column next to participant names) using RAND. Then sort the data by random number taking top half as control and bottom as intervention (for example)

- Use with large sample sizes
- Can suffer from ‘chance bias’
- However, avoid simple alternation. Do not use birth dates or alphabet (eg) as can introduce bias

National College for Teaching & Leadership

# Using RAND in Excel

**1.** **2.** **3.**

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Erica | 0.45251 | | John | 0.030918 | | John | 0.030918 | Control | |
| 2 | John | 0.030918 | | Marian | 0.041152 | | Marian | 0.041152 | Control | |
| 3 | Brian | 0.380752 | | Lisa | 0.308898 | | Lisa | 0.308898 | Control | |
| 4 | Peter | 0.998792 | | Richard | 0.360259 | | Richard | 0.360259 | Control | |
| 5 | Mohammed | 0.374822 | | Mohamm | 0.374822 | | Mohamm | 0.374822 | Intervention | |
| 6 | Lisa | 0.308898 | | Brian | 0.380752 | | Brian | 0.380752 | Intervention | |
| 7 | Richard | 0.360259 | | Erica | 0.45251 | | Erica | 0.45251 | Intervention | |
| 8 | Marian | 0.041152 | | Peter | 0.998792 | | Peter | 0.998792 | Intervention | |
| 9 | | | | | | | | | | |

# Types of randomisation and their limitations

**Stratified randomisation** (randomly allocate controlling for some variables (particularly participant characteristics))

- For example, by making sure there are equal numbers of males and females, children with SEN etc.
- Creating separate ‘blocks’ of data for different groups can help to ensure a balanced sample

- Be careful that you do not introduce another bias
- Some primary and nursery schools have naturally occurring forms of random stratification on entry; make use of these, or introduce into your school admissions structure

# Other forms (1)

## Pairwise randomisation

- Recruit two participants at a time and randomly allocate as you build up the sample

- Useful where there are ‘intervention’ slots (used in surgery and counselling)
- Similar issues to simple randomisation if sample is small

National College for Teaching & Leadership

# Other forms (2)

## Matched pair design (another form of between-subject design)

- A research design in its own right
- Participants randomly allocated after the ‘coupling’ (pairing) of participants with similar attributes such as age, height, interests etc
- Comes close to the advantages of a within-subject design in controlling for between-subject variation
- Time-consuming and never perfect

National College for
Teaching & Leadership

# Statistical analysis

# Why would you want to know this?

- You have invested a lot of time and effort into designing and implementing an RCT and you need to know what you have found
- You need to be able to talk about it in a way that other quantitative researchers will understand and respect
- You may want to publish your results
- Just as it is essential to use the same words that your interviewees used, when writing up some forms of qualitative research, so using the right statistic in the right way is important in quantitative research

# Statistical analysis - the key principles

- The type of data you have determines what test to use
- How your data is distributed is important because different tests are suitable for different distributions (parametric tests for normally distributed data and non-parametric for non-normally distributed data)
- There are also different tests for use with between-subject (independent samples) and within-subject (repeated measures) designs
- However, all tests produce the statistic p (e.g. $p = .001$) and therefore tell you whether there is a difference or not
- There are different reporting conventions for each test (and different styles within these conventions. . .)

# Types of statistics

- **Descriptive statistics** (mean (or median), standard deviation etc) – you will need to also include these somewhere in your report

- **Inferential statistics** (produces a probability level) – the p level tells you whether the results are likely to have occurred by chance (i.e. whether the results for control and intervention are essentially the same). The minimum usually considered acceptable is $p < .05$ (a five in one hundred probability the difference could have occurred by chance)

# Types of data

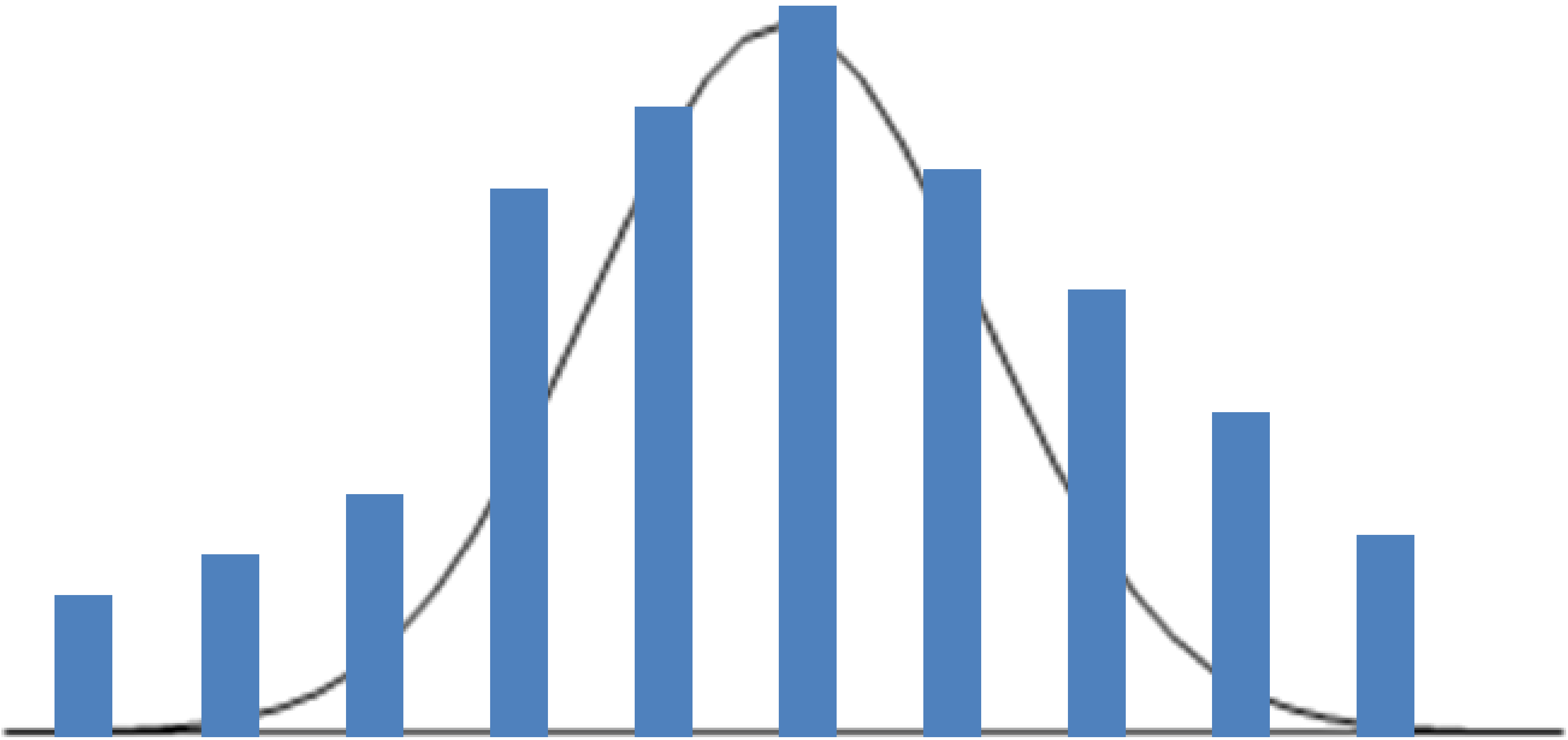
- **Interval** - data that comes from numerical scales in which the order is known as well as exact differences between values (e.g. overall test scores, height in centimetres etc)
- **Ordinal** - data in which only the order is known (e.g. rank ordering on a Likert scale)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Completely | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Not at all |

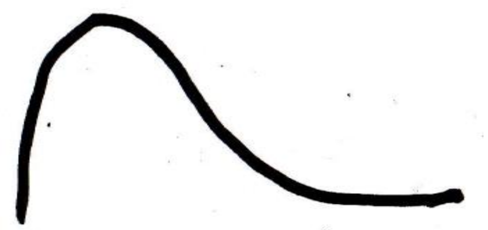
- **Categorical** – amounts of something that can be counted (e.g. 25 passes; 27 fails)

National College for Teaching & Leadership

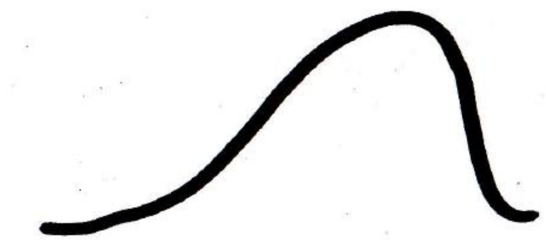
# Normal distribution



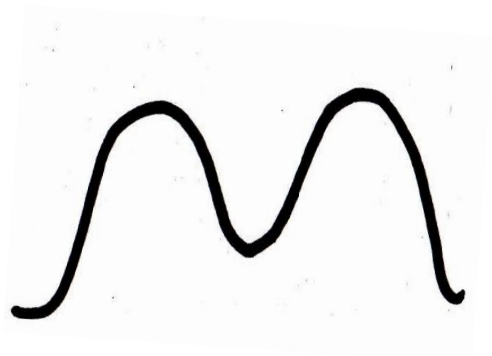
# Other distributions



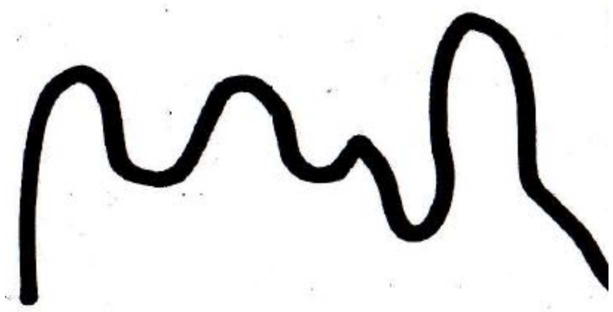
Positive skew



Negative skew



Bimodal

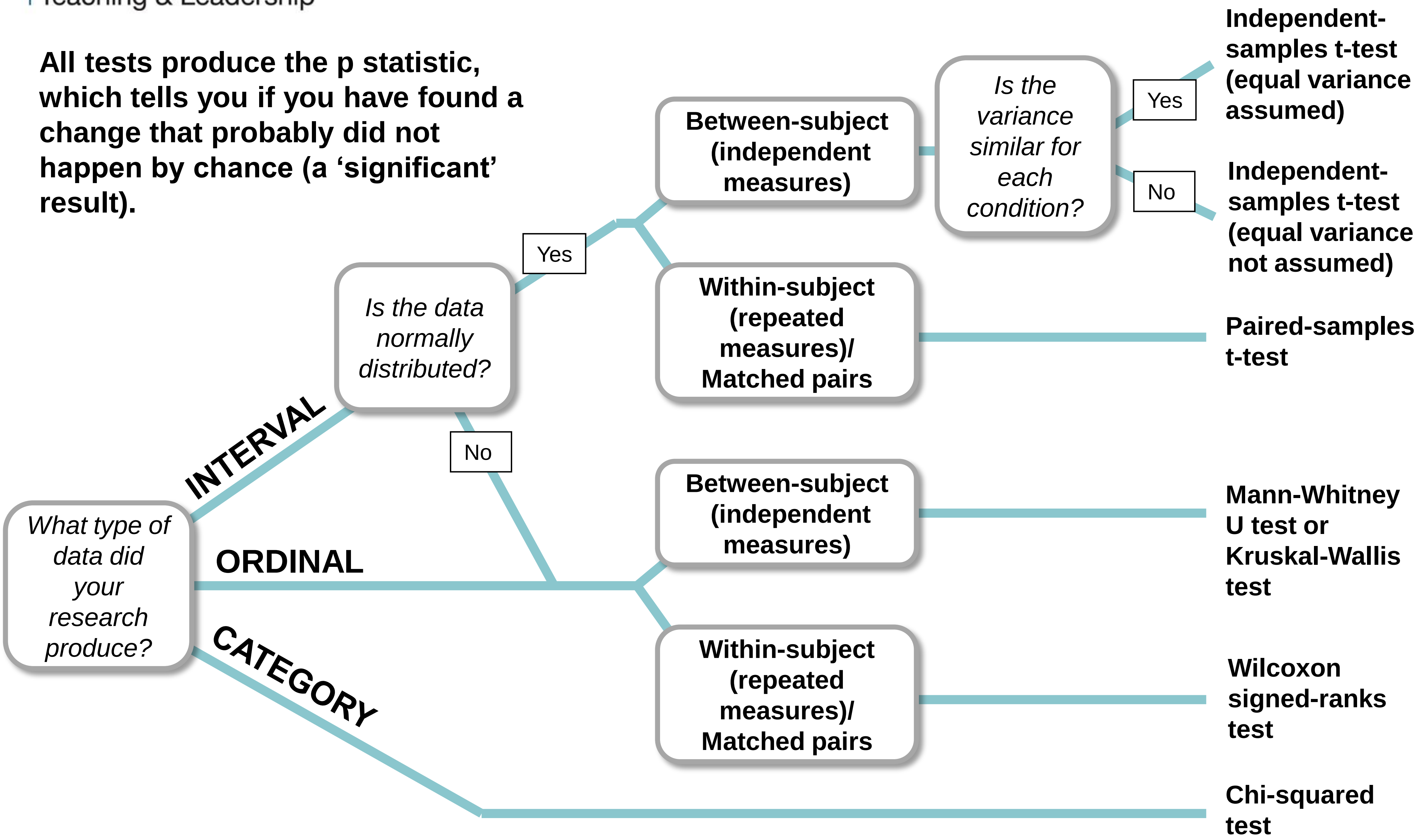


Multimodal

National College for Teaching & Leadership

# What test to use?

**All tests produce the p statistic, which tells you if you have found a change that probably did not happen by chance (a 'significant' result).**



National College for Teaching & Leadership

# Hypothesis testing

- You should decide in advance if your hypothesis is one-tailed (you think the scores will go one way) or two-tailed (you don't know which way they will go)
- And set a probability level (the threshold at which you will say that you have found something or not). This could be $p < .05$, $p < .01$ or even greater ($p < .001$). This is technically known as alpha (e.g. $\alpha = .05$)
- If you don't reach this threshold you accept your null hypothesis, if you cross it you reject your null hypothesis

National College for Teaching & Leadership

# Rejecting or accepting the null hypothesis

**For example with alpha = .05:**

**If p = .678, accept your null hypothesis:** Three lessons of group work does not improve maths attainment as measured by the correct number of column addition problems solved in a 10-minute test

**If p = .049, reject your null hypothesis and accept your experimental hypothesis:** Three lessons of group work improves maths attainment as measured by the correct number of column addition problems solved in a 10-minute test

# Practical activity 1

- In pairs, look at the data in the Practical Activities sheets
- For your sheets (to be agreed in the room) – identify:

  - The research design
  - The type of data
  - The direction of the hypothesis
  - Note the significant level set as the threshold (alpha)

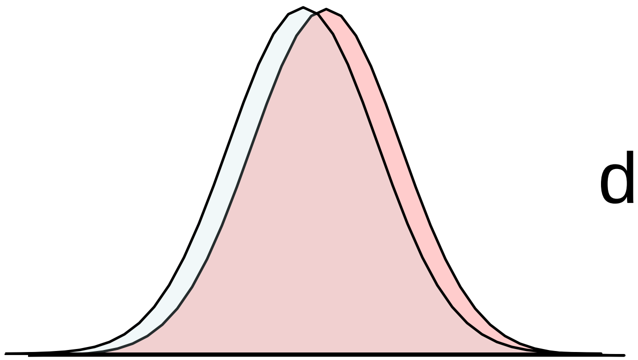
National College for Teaching & Leadership

# Cohen’s d

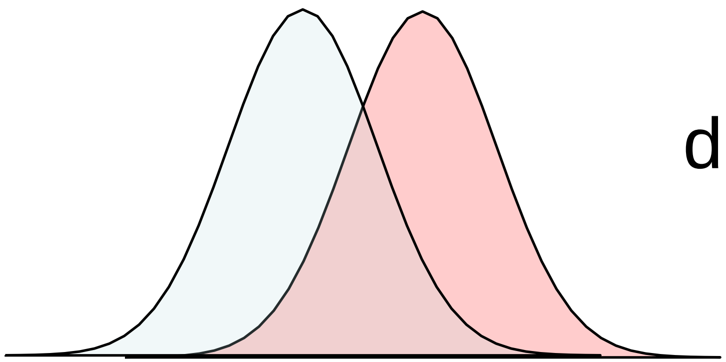
$$Effect\ size\ d = \frac{Intervention\ mean - Control\ mean}{pooled\ Standard\ Deviation}$$

National College for Teaching & Leadership

# For example



d = 0.1



d = 0.8

| Effect Size d | Percentage of pupils likely to have been clearly affected by the intervention, if the result is significant (percentage of non-overlap between distributions) |
|---|---|
| 1.3 | 65.3% |
| 1.2 | 62.2% |
| 1.1 | 58.9% |
| 1.0 | 55.4% |
| 0.9 | 51.6% |
| 0.8 | 47.4% |
| 0.7 | 43.0% |
| 0.6 | 38.2% |
| 0.5 | 33.0% |
| 0.4 | 27.4% |
| 0.3 | 21.3% |
| 0.2 | 14.7% |
| 0.1 | 7.7% |
| 0.0 | 0% |

# Reporting a t-test

- An independent samples t-test indicated a significant ($p = .001$ (one-tailed)) improvement in attainment for the pupils who were exposed to the OMS method ($M = 17.4$, $SD = 3.4$) compared to the control ($M = 10.3$, $SD = 2.4$). This represented a moderate effect size ($d = 0.4$).

# Reporting a Mann-Whitney test

- A Mann-Whitney U test indicated a significant difference ($p = 0.01$ (one-tailed) between the new behaviour management approach (Mdn = 5.3) and the school's current practice (Mdn = 1.2). The new approach appears to have a large effect on pupil behaviour ($r = 0.81$).

# Practical activity 2

- Pick a research design from sheets A, B, C or D and analyse the data

- Then write up the result on a piece of flipchart paper

# Different effect sizes are appropriate for different distributions and tests

| | **Between two conditions** | | | **Across three conditions** | |
|---|---|---|---|---|---|
| | Normally distributed interval data | Non-normal interval data and ordinal data | Category data | Normally distributed interval data | Non-normal interval data and ordinal data |
| | **d** | **r** | **w (Phi)** | **$n_p^2$ (partial eta squared)** | **W (Kendal's)** |
| Small | 0.20 | 0.10 | 0.10 | 0.01 | 0.20 |
| Medium | 0.50 | 0.30 | 0.30 | 0.06 | 0.40 |
| Large | 0.80 | 0.50 | 0.50 | 0.14 | 0.60 |

National College for Teaching & Leadership

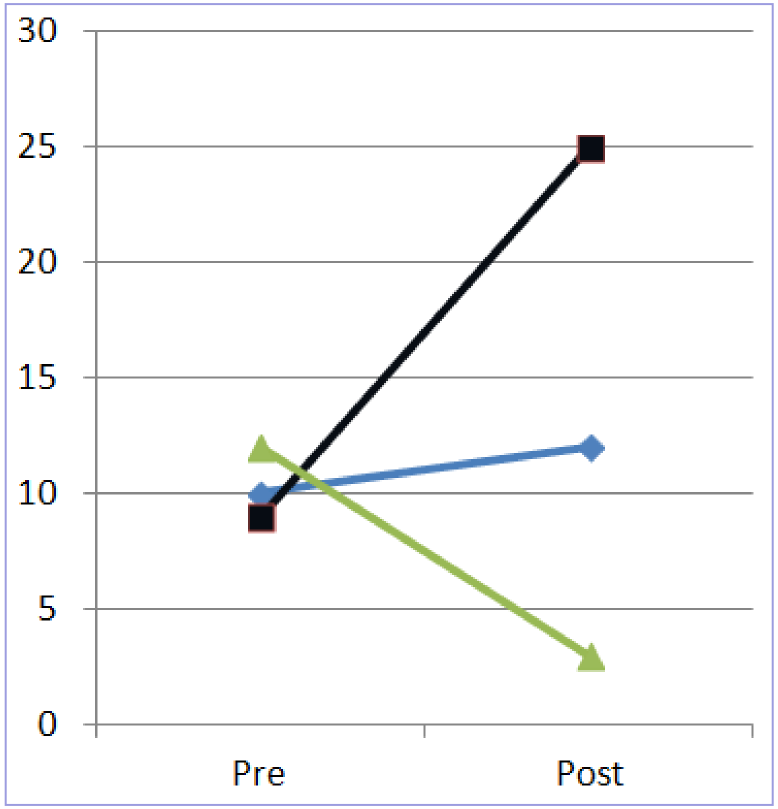
# Interpreting the effect size d (according to EEF) – useful for extended trials using standardised tests

| Months' progress | Effect Size from … | ... to |
|---|---|---|
| 0 | -0.01 | 0.01 |
| 1 | 0.02 | 0.09 |
| 2 | 0.10 | 0.18 |
| 3 | 0.19 | 0.26 |
| 4 | 0.27 | 0.35 |
| 5 | 0.36 | 0.44 |
| 6 | 0.45 | 0.52 |
| 7 | 0.53 | 0.61 |
| 8 | 0.62 | 0.69 |
| 9 | 0.70 | 0.78 |
| 10 | 0.79 | 0.87 |
| 11 | 0.88 | 0.95 |
| 12 | 0.96 | >1.0 |

Higgins, S., Kokotsaki, D. and Coe, R. (2012) The teaching and learning toolkit: technical appendices, Durham University: CEM.

# If you have pre- and post- test data

- If you can, use ANCOVA (which controls for variation in pre-test scores)
- Or use Gain Scores (post-test score minus pre-test score for each participant in your study) and put these into the analysis
- Note:  if you use Gain Scores your hypothesis is now related to progress not attainment



National College for Teaching & Leadership

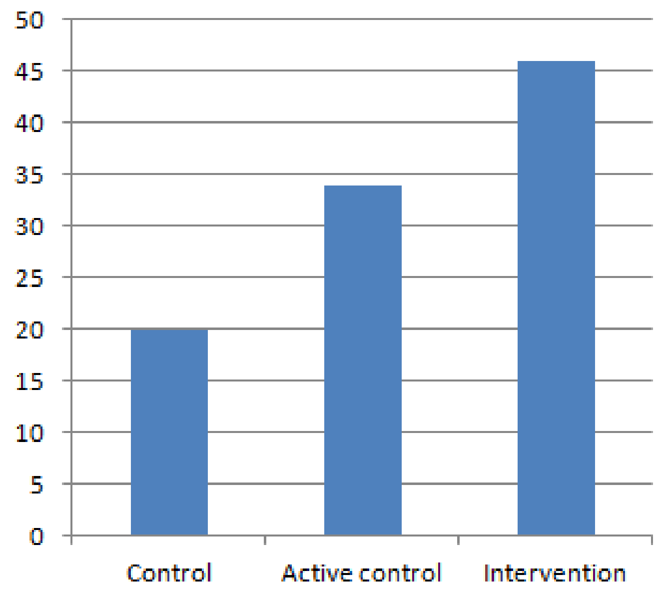
# Reporting a pre- post-test design where you have used ANCOVA or gain scores

- ANCOVA with pre-test scores as the covariate indicated that there was no difference between the control and intervention groups ($p = 0.76$ (one-tailed)). The table below shows pre- and post-test scores for both conditions.

- Using gain scores an independent sample t-test indicated that boys' progress following the consumption of the 'Yukky' drink (mean difference = 57.4, SD = 13.1) was significantly ($p = .01$ (two-tailed)) lower than girls' progress (mean difference = 78.4, SD = 14.5). This represented a small effect size ($d = 0.2$).

# If you have three conditions and a post-test design

- First, use an ANOVA to see if the change across all conditions is significant (if it is not you may have a family wise error so say so and be cautious in your interpretation)
- Then report the planned comparisons (condition 1 vs 2; condition 2 vs 3; condition 1 vs 3), reporting all results. Use a Bonferroni adjusted alpha. For example, if you have a three condition design, a = .05 now becomes α = .017 (i.e. p = .017 is now your cut-off)

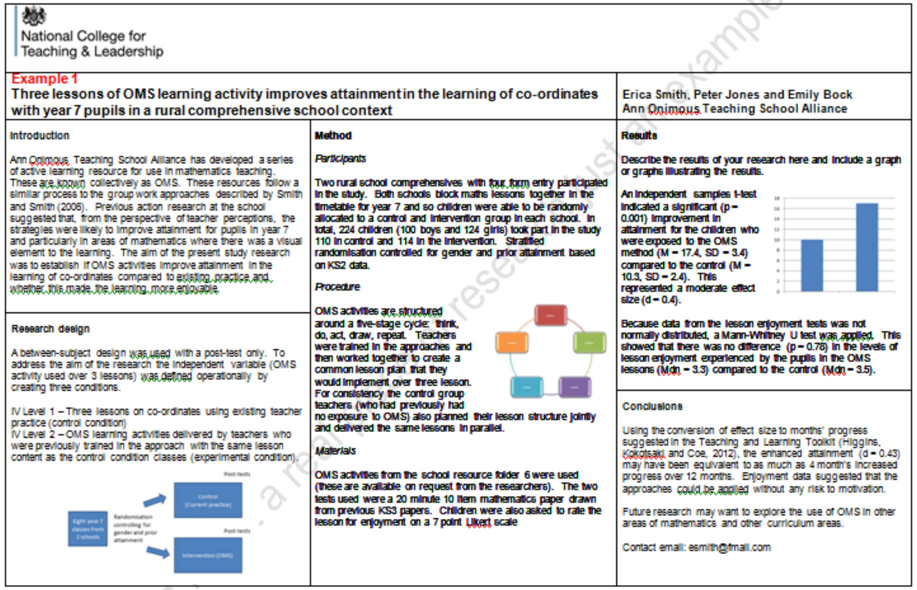


# Reporting the results from a three condition study

- Analysis used ANOVA with planned comparison. The initial ANOVA across all three conditions indicated no significant change. Pairwise contrasts were then conducted. Because of the use of multiple tests a more stringent significance level (known as a Bonferonni adjustment) was applied (0.17). The table below shows the effect sizes and levels of significance, comparing all three conditions to one another. There was no difference between the attainment of pupils who were marked in Green (M = 22.1, SD = 5.5) or Red (M = 21.5, SD = 4.6), compared to the control condition (M = 23.4, SD = 5.6).

# Look at some example conference poster results

- Read the results and conclusions sections in the example conference posters
- Notice the types of data that have been reported and how this has been done
- In some cases, months' gain has been discussed in the conclusions



National College for
Teaching & Leadership

Example 1

**Three lessons of OMS learning activity improves attainment in the learning of co-ordinates with year 7 pupils in a rural comprehensive school context**

Erica Smith, Peter Jones and Emily Bock
Ann Onimous Teaching School Alliance

**Introduction**

Ann Onimous Teaching School Alliance has developed a series of active learning resource for use in mathematics teaching. These are known collectively as OMS. These resources follow a similar process to the group work approaches described by Smith and Smith (2006). Previous action research at the school suggested that, from the perspective of teacher perceptions, the strategies were likely to improve attainment for pupils in year 7 and particularly in areas of mathematics where there was a visual element to the learning. The aim of the present study research was to establish if OMS activities improve attainment in the learning of co-ordinates compared to existing practice and whether this made the learning more enjoyable.

**Research design**

A between-subject design was used with a post-test only. To address the aim of the research the independent variable (OMS activity used over 3 lessons) was defined operationally by creating three conditions.

IV Level 1 – Three lessons on co-ordinates using existing teacher practice (control condition)
IV Level 2 – OMS learning activities delivered by teachers who were previously trained in the approach with the same lesson content as the control condition classes (experimental condition).



**Method**

*Participants*

Two rural school comprehensives with four form entry participated in the study. Both schools block maths lessons together in the timetable for year 7 and so children were able to be randomly allocated to a control and intervention group in each school. In total, 224 children (100 boys and 124 girls) took part in the study 110 in control and 114 in the intervention. Stratified randomisation controlled for gender and prior attainment based on KS2 data.

*Procedure*

OMS activities are structured around a five-stage cycle: think, do, act, draw, repeat. Teachers were trained in the approaches and then worked together to create a common lesson plan that they would implement over three lesson. For consistency the control group teachers (who had previously had no exposure to OMS) also planned their lesson structure jointly and delivered the same lessons in parallel.

*Materials*

OMS activities from the school resource folder 6 were used (these are available on request from the researchers). The two tests used were a 20 minute 10 item mathematics paper drawn from previous KS3 papers. Children were also asked to rate the lesson for enjoyment on a 7 point Likert scale

**Results**

Describe the results of your research here and include a graph or graphs illustrating the results.

An independent samples t-test indicated a significant ($p = 0.001$) improvement in attainment for the children who were exposed to the OMS method ($M = 17.4$, $SD = 3.4$) compared to the control ($M = 10.3$, $SD = 2.4$). This represented a moderate effect size ($d = 0.4$).

Because data from the lesson enjoyment tests was not normally distributed, a Mann-Whitney U test was applied. This showed that there was no difference ($p = 0.78$) in the levels of lesson enjoyment experienced by the pupils in the OMS lessons ($Mdn = 3.3$) compared to the control ($Mdn = 3.5$).

**Conclusions**

Using the conversion of effect size to months' progress suggested in the Teaching and Learning Toolkit (Higgins, Kokotsaki and Coe, 2012), the enhanced attainment ($d = 0.43$) may have been equivalent to as much as 4 month's increased progress over 12 months. Enjoyment data suggested that the approaches could be applied without any risk to motivation.

Future research may want to explore the use of OMS in other areas of mathematics and other curriculum areas.

Contact email: esmith@fmail.com

- Notice that where a parametric test is used, the mean, standard deviation and d are given, but where the test was non-parametric the median and r are reported

National College for
Teaching & Leadership

# Practical activity 3

- Pick a research design from either sheets E or F and analyse the data
- If you use F you could use ANCOVA or gain scores
- Then write up the result on a piece of flipchart paper

# Interpreting results – common points to remember

- If your control condition was a form of ‘existing practice’, then a non-significant result (e.g. $p > 0.05$) means that the intervention was equal to existing practice (i.e. you have identified an alternative treatment)
- If you have a non-significant result then you cannot claim that an effect size exists
- Remember that populations make a difference and avoid generalising beyond the particular group and context that your experiment used
- Name the test you used and, as a minimum, the significance (p) and the effect size

National College for Teaching & Leadership

# Writing up your research using a conference poster style

National College for Teaching & Leadership

HANDOUT 4

Example 1

## Three lessons of OMS learning activity improve attainment in the learning of co-ordinates with Year 7 pupils in a rural comprehensive school context

Erica Smith, Peter Jones and Emily Bock
Ann Onimous Teaching School Alliance



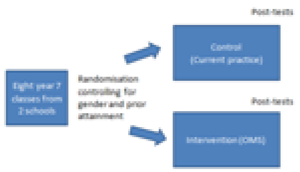
### Introduction

Ann Onimous Teaching School Alliance has developed a series of active learning resources for use in mathematics teaching. These are known collectively as OMS. These resources follow a similar process to the group work approaches described by Smith and Smith (2006). Previous action research at the school suggested that, from the perspective of teacher perceptions, the strategies were likely to improve attainment for pupils in Year 7 and particularly in areas of mathematics where there was a visual element to the learning. The aim of the present study research was to establish if OMS activities improved attainment in the learning of co-ordinates compared to existing practice and whether this made the learning more enjoyable. The research also aimed to establish if there was an improvement in Thinking Mind scores (Dweep, 1997).

### Research design

A between-subject design was used with a post-test only. To address the aims of the research the independent variable (OMS activity used over three lessons) was defined operationally by creating two conditions.

IV Level 1 – Three lessons on co-ordinates using existing teacher practice (control condition)
IV Level 2 – OMS learning activities delivered by teachers who were previously trained in the approach with the same lesson content as the control condition classes (experimental condition).




### Method

*Participants*

Two rural school comprehensives with four-form entry participated in the study. Both schools block maths lessons together in the timetable for Year 7 and so pupils were able to be randomly allocated to a control and intervention group in each school. In total, 224 pupils (100 boys and 124 girls) took part in the study – 110 in control and 114 in the intervention. Stratified randomisation controlled for gender and prior attainment based on KS2 data.

*Procedure*

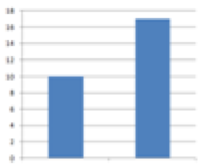
OMS activities are structured around a five-stage cycle: think, do, act, draw, repeat. Teachers were trained in the approaches and then worked together to create a common lesson plan that they would implement over three lessons. For consistency, the control group teachers (who had previously had no exposure to OMS) also planned their lesson structure jointly and delivered the same lessons in parallel.

*Materials*

OMS activities from the school resource folder 6 were used (these are available on request from the researchers). The two tests used were a 20-minute 10-item mathematics paper drawn from previous KS3 papers. Children were also asked to rate the lesson for enjoyment on a 7-point Likert scale; and complete questions 7–10 from the Thinking Mind questionnaire (Dweep, 1997).

### Results

An independent samples t-test indicated a significant ($p = .001$ (one-tailed)) improvement in attainment for the pupils who were exposed to the OMS method ($M = 17.4$, $SD = 3.4$) compared to the control ($M = 10.3$, $SD = 2.4$). This represented a moderate effect size ($d = 0.4$).



Because data from the lesson enjoyment tests and Thinking Mind questionnaire was not normally distributed, a Mann-Whitney U test was applied. This showed that there was no difference ($p = .78$ (two-tailed)) in the levels of lesson enjoyment experienced by the pupils in the OMS lessons ($Mdn = 3.3$) compared to the control ($Mdn = 3.5$). However, there was a significance difference ($p = .001$ (two-tailed)) between thinking mindset scores for the intervention group ($Mdn = 18.9$) compared to the control ($Mdn = 9.4$), a large effect ($r = 0.81$).

### Conclusions

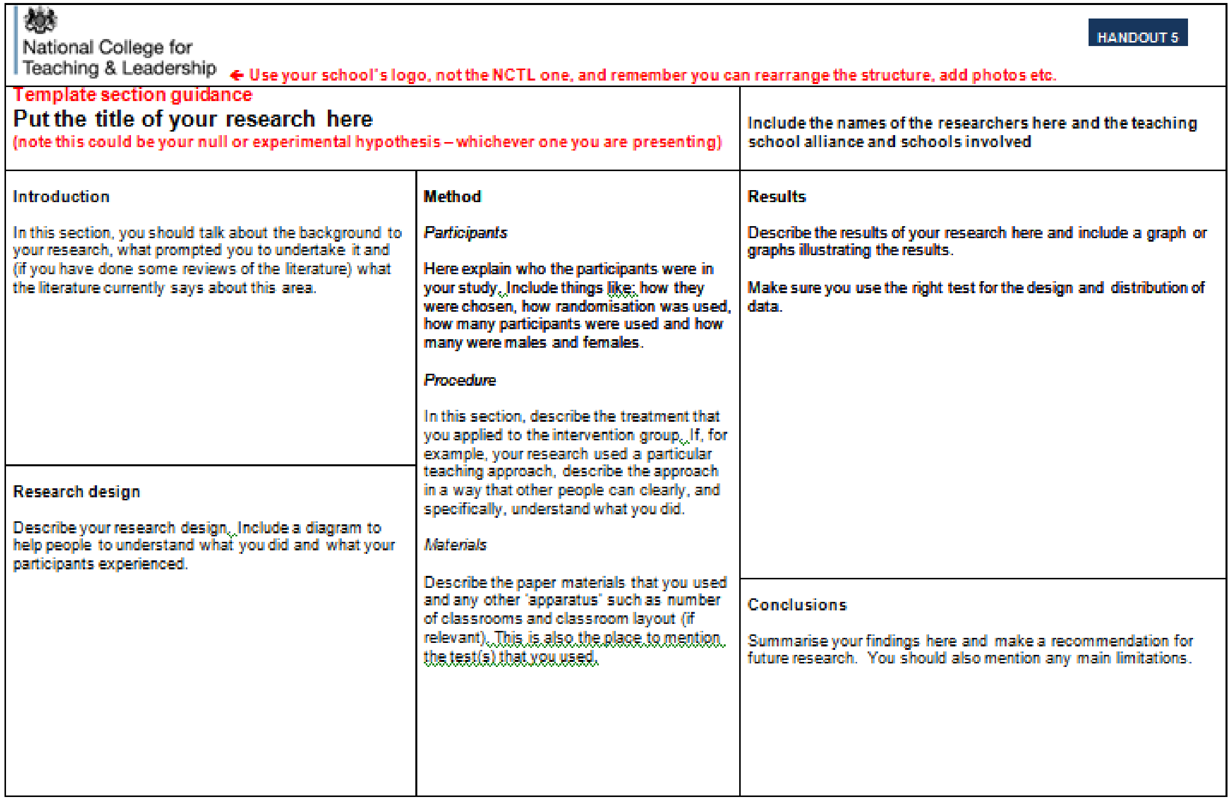
Using the conversion of effect size to months' progress suggested in the *Teaching and Learning Toolkit* (Higgins, Kokotsaki and Coe, 2012), the enhanced attainment ($d = 0.43$) may have been equivalent to as much as four months' increased progress over 12 months. Enjoyment data suggested that the approaches could be applied without any risk to motivation. Children who experienced the OMS lessons also showed a significant and large enhancement in their Thinking Mind.

Future research may want to explore the use of OMS in other areas of mathematics and other curriculum areas.
Contact email: esmith@fmail.com

1

- Introduction
- Research design
- Method (participants, procedure, materials)
- Results
- Conclusions

# Tips for what to include in each section of your report



National College for Teaching & Leadership ← Use your school's logo, not the NCTL one, and remember you can rearrange the structure, add photos etc.

HANDOUT 5

Template section guidance

**Put the title of your research here**
(note this could be your null or experimental hypothesis – whichever one you are presenting)

Include the names of the researchers here and the teaching school alliance and schools involved

**Introduction**

In this section, you should talk about the background to your research, what prompted you to undertake it and (if you have done some reviews of the literature) what the literature currently says about this area.

**Research design**

Describe your research design. Include a diagram to help people to understand what you did and what your participants experienced.

**Method**

*Participants*

Here explain who the participants were in your study. Include things like: how they were chosen, how randomisation was used, how many participants were used and how many were males and females.

*Procedure*

In this section, describe the treatment that you applied to the intervention group. If, for example, your research used a particular teaching approach, describe the approach in a way that other people can clearly, and specifically, understand what you did.

*Materials*

Describe the paper materials that you used and any other 'apparatus' such as number of classrooms and classroom layout (if relevant). This is also the place to mention the test(s) that you used.

**Results**

Describe the results of your research here and include a graph or graphs illustrating the results.

Make sure you use the right test for the design and distribution of data.

**Conclusions**

Summarise your findings here and make a recommendation for future research. You should also mention any main limitations.

- Look at the guidance template and consider: a) in the light of your own research and, b) what you need to/or might need to write up
- We will talk though each section first

# Publishing your poster – copyright and permissions

- You must seek written permission from the creator/owner to use any existing images or photographs
- Quotes or text from other sources must be referenced (you may also need to seek permission, particularly if you are changing the context)
- If you do not seek permissions or reference sources, you could be sued by the creator/owner

National College for Teaching & Leadership

# Helpline and support

- Telephone: 0118 902 1919 (also for help with randomisation)
- Available Monday to Friday (10am – 3pm) with answerphone
- Email address: ctg@cfbt.com
- Or specifically to do with technical aspects of your research or analysis: rchurches@cfbt.com