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
• This event is for those who have attended no RDNEs or just RDNE 1 (will recap RDNE 1, 2 and 3 content)
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?
Why randomised controlled trials?

- 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)
Reminding ourselves of the basic approach

Randomised

(Cabinet Office, 2012)

Population is split into 2 groups by random lot

Controlled

Outcomes for both groups are measured

TRIAL

INTERVENTION

CONTROL

© Crown copyright 2015

Closing the Gap: Test and Learn
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)?
• 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
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)
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 \(\rightarrow\) intervention or intervention \(\rightarrow\) 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)
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

• Does group work improve column addition problem-solving ability?
• 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**
• Is your measure valid and reliable? Imagine the centre of the target is your research aim – does your DV relate to your IV?

Valid and reliable

Reliable but not valid

Valid but not reliable

Not valid and not reliable
Back to the template

- Research topic/area
- Research aim(s)
- Research question(s)
- Research design
- Method

What is important about this?

How could you measure this?

What is your hypothesis?

Who, what and how?

Is a quantitative method the best way to do this?
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
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 (e.g., performance improvement), usually dealt with by counterbalancing
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)
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
For example

<table>
<thead>
<tr>
<th>Effect Size d</th>
<th>Percentage of pupils likely to have been clearly affected by the intervention, if the result is significant (percentage of non-overlap between distributions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3</td>
<td>65.3%</td>
</tr>
<tr>
<td>1.2</td>
<td>62.2%</td>
</tr>
<tr>
<td>1.1</td>
<td>58.9%</td>
</tr>
<tr>
<td>1.0</td>
<td>55.4%</td>
</tr>
<tr>
<td>0.9</td>
<td>51.6%</td>
</tr>
<tr>
<td>0.8</td>
<td>47.4%</td>
</tr>
<tr>
<td>0.7</td>
<td>43.0%</td>
</tr>
<tr>
<td>0.6</td>
<td>38.2%</td>
</tr>
<tr>
<td>0.5</td>
<td>33.0%</td>
</tr>
<tr>
<td>0.4</td>
<td>27.4%</td>
</tr>
<tr>
<td>0.3</td>
<td>21.3%</td>
</tr>
<tr>
<td>0.2</td>
<td>14.7%</td>
</tr>
<tr>
<td>0.1</td>
<td>7.7%</td>
</tr>
<tr>
<td>0.0</td>
<td>0%</td>
</tr>
</tbody>
</table>
• Effect size – difference in mean, taking into account the spread of scores (or variance). One way to do this is with Cohen’s $d$.

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

• 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.
$p$ (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))

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.
• Some designs could be carried out where a large group of children are blocked in the timetable (e.g. 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 (e.g. in a single lesson design where children were randomly allocated to different tasks in the same lesson; or two parallel classes)
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.
• 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…
• 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

• Read

• Discuss

• Present back in groups
Randomisation and EXCEL randomisation activity
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)
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
Using RAND in Excel

1.  
2.  
3.  

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Erica</td>
<td>0.45251</td>
<td></td>
<td></td>
<td>John</td>
<td>0.030918</td>
<td></td>
<td>John</td>
<td>0.030918</td>
<td>Control</td>
</tr>
<tr>
<td>2</td>
<td>John</td>
<td>0.030918</td>
<td></td>
<td></td>
<td>Marian</td>
<td>0.041152</td>
<td></td>
<td>Marian</td>
<td>0.041152</td>
<td>Control</td>
</tr>
<tr>
<td>3</td>
<td>Brian</td>
<td>0.380752</td>
<td></td>
<td></td>
<td>Lisa</td>
<td>0.308898</td>
<td></td>
<td>Lisa</td>
<td>0.308898</td>
<td>Control</td>
</tr>
<tr>
<td>4</td>
<td>Peter</td>
<td>0.998792</td>
<td></td>
<td></td>
<td>Richard</td>
<td>0.360259</td>
<td></td>
<td>Richard</td>
<td>0.360259</td>
<td>Control</td>
</tr>
<tr>
<td>5</td>
<td>Mohammed</td>
<td>0.374822</td>
<td></td>
<td></td>
<td>Mohammed</td>
<td>0.374822</td>
<td></td>
<td>Mohammed</td>
<td>0.374822</td>
<td>Intervention</td>
</tr>
<tr>
<td>6</td>
<td>Lisa</td>
<td>0.308898</td>
<td></td>
<td></td>
<td>Brian</td>
<td>0.380752</td>
<td></td>
<td>Brian</td>
<td>0.380752</td>
<td>Intervention</td>
</tr>
<tr>
<td>7</td>
<td>Richard</td>
<td>0.360259</td>
<td></td>
<td></td>
<td>Erica</td>
<td>0.45251</td>
<td></td>
<td>Erica</td>
<td>0.45251</td>
<td>Intervention</td>
</tr>
<tr>
<td>8</td>
<td>Marian</td>
<td>0.041152</td>
<td></td>
<td></td>
<td>Peter</td>
<td>0.998792</td>
<td></td>
<td>Peter</td>
<td>0.998792</td>
<td>Intervention</td>
</tr>
</tbody>
</table>
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
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
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
Statistical analysis
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.
• 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. . .)
• **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)

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

Positive skew

Negative skew

Bimodal

Multimodal
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).

- **Interval**
  - *What type of data did your research produce?*
    - **Ordinal**
      - *Is the data normally distributed?*
        - Yes
          - Between-subject (independent measures)
        - No
          - Within-subject (repeated measures)/Matched pairs
    - **Category**
      - *Is the variance similar for each condition?*
        - Yes
          - Independent-samples \( t \)-test (equal variance assumed)
        - No
          - Independent-samples \( t \)-test (equal variance not assumed)
- Paired-samples \( t \)-test
- Mann-Whitney \( U \) test or Kruskal-Wallis test
- Wilcoxon signed-ranks test
- Chi-squared test

---

© Crown copyright 2015
• 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.
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.
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)
Cohen’s $d$

\[
\text{Effect size } d = \frac{\text{Intervention mean} - \text{Control mean}}{\text{pooled Standard Deviation}}
\]
For example

For example

<table>
<thead>
<tr>
<th>Effect Size d</th>
<th>Percentage of pupils likely to have been clearly affected by the intervention, if the result is significant (percentage of non-overlap between distributions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3</td>
<td>65.3%</td>
</tr>
<tr>
<td>1.2</td>
<td>62.2%</td>
</tr>
<tr>
<td>1.1</td>
<td>58.9%</td>
</tr>
<tr>
<td>1.0</td>
<td>55.4%</td>
</tr>
<tr>
<td>0.9</td>
<td>51.6%</td>
</tr>
<tr>
<td>0.8</td>
<td>47.4%</td>
</tr>
<tr>
<td>0.7</td>
<td>43.0%</td>
</tr>
<tr>
<td>0.6</td>
<td>38.2%</td>
</tr>
<tr>
<td>0.5</td>
<td>33.0%</td>
</tr>
<tr>
<td>0.4</td>
<td>27.4%</td>
</tr>
<tr>
<td>0.3</td>
<td>21.3%</td>
</tr>
<tr>
<td>0.2</td>
<td>14.7%</td>
</tr>
<tr>
<td>0.1</td>
<td>7.7%</td>
</tr>
<tr>
<td>0.0</td>
<td>0%</td>
</tr>
</tbody>
</table>
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).
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 ($r = 0.81$).
• 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

<table>
<thead>
<tr>
<th></th>
<th>Between two conditions</th>
<th>Across three conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normally distributed</td>
<td>Normally distributed</td>
</tr>
<tr>
<td></td>
<td>interval data</td>
<td>interval data</td>
</tr>
<tr>
<td></td>
<td>Non-normal interval</td>
<td>Non-normal interval</td>
</tr>
<tr>
<td></td>
<td>data and ordinal data</td>
<td>data and ordinal data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0.20</td>
<td>0.01</td>
</tr>
<tr>
<td>r</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>w (Phi)</td>
<td>0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>n_p^2 (partial eta</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>squared)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W (Kendal’s)</td>
<td>0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Months’ progress</th>
<th>Effect Size from...</th>
<th>... to</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.19</td>
<td>0.26</td>
</tr>
<tr>
<td>4</td>
<td>0.27</td>
<td>0.35</td>
</tr>
<tr>
<td>5</td>
<td>0.36</td>
<td>0.44</td>
</tr>
<tr>
<td>6</td>
<td>0.45</td>
<td>0.52</td>
</tr>
<tr>
<td>7</td>
<td>0.53</td>
<td>0.61</td>
</tr>
<tr>
<td>8</td>
<td>0.62</td>
<td>0.69</td>
</tr>
<tr>
<td>9</td>
<td>0.70</td>
<td>0.78</td>
</tr>
<tr>
<td>10</td>
<td>0.79</td>
<td>0.87</td>
</tr>
<tr>
<td>11</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>12</td>
<td>0.96</td>
<td>&gt;1.0</td>
</tr>
</tbody>
</table>

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
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 ($d = 0.2$).
• 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 $a = .017$ (i.e. $p = .017$ is now your cut-off)
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 Bonferroni 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).
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.

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.
• 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
Writing up your research using a conference poster style

**Example 1**

<table>
<thead>
<tr>
<th>Three lessons of OMS learning activity improve attainment in the learning of co-ordinates with Year 7 pupils in a rural comprehensive school context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Arm Chrysalis Teaching School Alliance has developed a series of active learning resources for use in mathematics teaching. These resources follow a similar process to the group work approaches described by Smith and Smith (2006). Previous ad hoc 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 Mindset scores (Dweck, 1997).</td>
</tr>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
</tr>
<tr>
<td>Two rural school comprehensive with four-form entry participants in the study, both schools block teach 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. Initially, 224 pupils (103 boys and 121 girls) took part in the study. In total, 110 of the sample were in the intervention <em>stimulated randomisation</em> condition for co-ordinates and prior attainment based on KS2 data.</td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
</tr>
<tr>
<td>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 utilised their lesson structure jointly and delivered the same lesson in parallel.</td>
</tr>
<tr>
<td><strong>Materials</strong></td>
</tr>
<tr>
<td>OMS activities from the school resource folder (OMS learning activities delivered by teachers who were previously trained in the approach with the same lesson context as the control condition classes (experimental condition).</td>
</tr>
<tr>
<td><strong>Results</strong></td>
</tr>
<tr>
<td>An independent samples t-test indicated a significant difference (p = .001; two-tailed) between thinking mindset scores for the intervention group (Mn = 18.3) compared to the control group (Mn = 16.4), a large effect (r = 0.4).</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Drawing data from the lesson observation tools and Thinking Mindset questionnaire was not normally distributed. A Mann-Whitney U test was applied. This showed that there was no difference (p = .76; two-tailed) in the levels of lesson enjoyment experienced by the pupils in the OMS lesson (M = 5.9) compared to the control (M = 5.3). However, there was a significant difference (p = .001; two-tailed) between thinking mindset scores for the intervention group (Mn = 18.3) compared to the control group (Mn = 16.4), a large effect (r = 0.4).</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
</tr>
<tr>
<td>Using the conversion of effect sizes to monthly progress suggested in the Teaching and Learning Toolkit (Dweck, Johnson, and Coe, 2012), the enhanced attainment (r = 0.44) 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 Mindsets.</td>
</tr>
<tr>
<td>Future research may want to explore the use of OMS in other areas of mathematics and other curriculum areas.</td>
</tr>
<tr>
<td>Contact email: <a href="mailto:asmith@gmail.com">asmith@gmail.com</a></td>
</tr>
</tbody>
</table>

- Introduction
- Research design
- Method (participants, procedure, materials)
- Results
- Conclusions
Tips for what to include in each section of your report

- 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

<table>
<thead>
<tr>
<th>Template section guidance</th>
<th>Include the names of the researchers here and the teaching school alliance and schools involved</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Put the title of your research here</strong></td>
<td>(note this could be your null or experimental hypothesis – whichever you are presenting)</td>
</tr>
</tbody>
</table>

**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.

**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.

**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.

**Research design**

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

**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 tests that you used.

**Conclusions**

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