

Innovation evaluation handbook

Institute for Effective Education
September 2017

What do we mean by an innovation evaluation?

The Research Schools Network is a joint initiative of the Institute for Effective Education (IEE) and the Education Endowment Foundation (EEF) and has the overall goal of improving the attainment of pupils by increasing the use of evidence-based teaching and learning practices. As part of this work, the IEE is providing innovation evaluation grants. The goal of the innovation evaluation fund is to increase the evidence base of what works in education by conducting evaluations of innovations in teaching and learning approaches, communicating the findings across the Research Schools Network and beyond, and encouraging applications for larger, rigorous evaluations of promising approaches.

The innovations that are supported by this fund are either approaches that teachers or schools have developed themselves or they are adaptations of approaches that have been proven to work elsewhere, which the teachers have altered to work with their particular pupils. The funded evaluations will be *pilot studies* of innovations, not large-scale *randomised controlled trials*. Nevertheless, there should be a comparison of the achievement of pupils who have not received the innovation with those who have, which will indicate whether or not the intervention has potential to improve outcomes for children.

The Innovation Evaluation Handbook

The purpose of this document is to support schools in planning and carrying out evaluations of practice. It is designed to be used while writing an application for an innovation evaluation grant, but the handbook could also be used to support the planning and execution of robust evaluations in schools. If you would like more information about the process of applying for an innovation evaluation grant, please see the [call for applications](#) for innovation evaluation grants. If you would like more information about writing up your evaluation, please see [Writing up your innovation evaluation report](#).

We have tried to avoid jargon in this handbook and have included a glossary in the Appendix.

Please note, in this document the word “school” is used to refer to any education setting for children aged 2-18 years, including early years settings, special schools, alternative provision and colleges.

Steps in conducting an evaluation

In order to conduct a robust *impact evaluation*, you should follow this process:

- Introduction
 - a. Describe the problem
 - b. Review existing evidence
 - c. Propose a potential solution
 - d. Present your research questions/hypotheses
- Method (planning the evaluation)
 - a. Sample
 - b. Assignment to condition
 - c. Innovation
 - d. Outcome measures
 - e. Process evaluation
 - f. Analyses
- Carry out the evaluation
- Results
 - a. Outcome findings
 - b. Process analysis
- Discussion/conclusion
 - a. Interpret your findings
 - b. Limitations
 - c. Implications for practice
 - d. Implications for further evaluation
 - e. Conclusions
- Communicate your findings

Introduction

a) Describe the problem

It is likely that your evaluation originates from an issue in your school or across a group of schools you work with. This issue could be identified through a range of sources eg, school level data, staff observations, pupil feedback, etc. Describe the challenge that your school or group of schools face.

b) Review existing evidence

Read about your identified problem to establish what is already known about the issue, how existing research evidence suggests that this issue could be addressed and whether there are any existing solutions with a good evidence base of effectiveness (if there is a well-evidenced solution already available, you can stop here! You should monitor impact of this solution in your school, but a formal evaluation will not be necessary).

You might draw on ratings of effective practices from the [EEF Teaching and Learning Toolkit](#), the IEE's [Evidence4Impact](#) or other websites that review research that has been conducted on

teaching and learning practices. Some of these are listed in the Resources section in the Appendix.

c) Propose your solution

Use your review of the evidence to decide on a potential solution to address your problem. Describe the innovation you will evaluate.

What existing research evidence supports your view that this innovation will be effective? This might be a review of similar practices within the research literature; you could consider within-school data of the impact of a similar approach being used in the past; or you might have chosen an evidence-based intervention that you are adapting for use with your particular pupils. Create a **logic model**, explaining the mechanisms by which you think it works.

d) Specific research question or hypothesis

Write the specific research question that you would like the evaluation to address. Having a well-defined research question is vital in planning and carrying out a coherent evaluation. The research question should include information about the practice that will be evaluated, for how long it will be delivered, the outcomes to be measured and the people who will experience the new practice. The research question could take the form:

What impact does [what practice?] delivered [over how long?] have on [what outcome?] for [whom]?

For example:

What impact does **Toddler Talk** delivered **for 10 minutes a day, 4 days a week for 4 months** have on **the expressive vocabulary** of **2 year olds whose home language is not English**?

You may also like to make a specific prediction about what you expect to happen. This should include the comparisons being made (eg, after 4 months, the average expressive vocabulary of 2 year olds who participate in Toddler Talk will be larger than the average expressive vocabulary of 2 year olds in the control condition) and should be based on what existing research suggests would be a likely outcome.

Method

You now need to plan an evaluation that will answer your research question. It is important to make decisions about the sample, recruitment, design, measures, intervention and data analysis before starting the evaluation so the research is scientifically valid (making decisions as you go along is likely to lead to biased results). Involve other stakeholders in your decision making, and write your plans clearly so you and everyone else involved understands what will happen. Create copies of research plans to be signed by all staff and any researchers involved (projects which receive innovation evaluation grants will be supported in developing their research plan by the IEE). It is also important to ensure you have buy-in from the leadership teams of all schools involved in the evaluation.

a) Sample and recruiting the participants

Consider who will participate in your evaluation. The **sample** should be representative of the group referred to in your research question. Decide on the required characteristics of **participants** in your sample (eg, year group or age, current attainment, home language, gender, etc); these required characteristics are called the **inclusion criteria**.

Decide how you will recruit participants to your study. If you plan to recruit pupils beyond your school, decide how and when you will approach other teachers/schools to share the opportunity to participate. Ensure you have buy-in from the school(s) leadership team(s), and conduct meetings with staff from the schools involved to be sure that everyone is clear about what is going to happen.

Decide whether you need consent for pupils to participate, who would give this consent and what sort of consent should be sought. For pupils under the age of 16, parental consent should be obtained before the evaluation begins. Student consent should also be sought when participants are over 16 years old. If consent would not be requested for pupils to participate in the innovation as part of normal classroom practice, it may only be necessary to ask for consent for pupil data to be included in data analysis (ie, all pupils will participate in the innovation, but parents can decide whether that child's data will be part of the analysis). If the innovation is a change in practice that a teacher might implement and is not very obtrusive, opt-out consent (ie, parents respond if they do *not* want their child to participate in the evaluation) rather than opt-in consent can increase participation rates. See the Appendix for a sample opt-out letter of consent. Whatever form of consent you choose, it is your responsibility to ensure that procedures and assessments are not potentially harmful.

b) Assignment to condition

Pupils usually make progress over time; we want to find out whether your innovation leads to improved progress compared to current practice. In order to establish the impact of your innovation, it is important to compare outcomes for people who experience the new practice with similar people who do not. This protects against the risk of typical progress (eg, phonetic awareness increases during Year 1) or general patterns (eg, a year-on-year national, regional or school-wide improvement in the phonetic awareness of Year 1 pupils) being misinterpreted as an effect caused by the innovation (eg, concluding that the new phonics programme improved the phonetic awareness of Year 1 pupils who received the intervention).

Condition refers to the group participants are in: they may be in the **intervention group** (sometimes called the experimental group) or the **control group** (sometimes called the comparison group). The intervention group will experience the practice you are evaluating, while the control group will experience whatever their teacher or school would usually deliver. The intervention and control group should be as similar as possible to each other at the beginning of your study. **Assignment to condition** refers to the way in which participants are allocated to different conditions.

You should assign participants to condition at the level at which the innovation is delivered. For example, if it is delivered to individuals (eg, a one-to-one tutoring programme), then you can assign individuals to control and intervention groups. If the innovation is delivered to whole classes (eg, a new maths curriculum) then it is better to assign classes to control and intervention groups. If the innovation is a whole-school intervention (eg, a school-wide healthy eating

programme), then it is best to assign schools.

Random assignment of participants to the control or intervention group is preferred but not necessary. You can assign to intervention and control groups by using a random number generator; there is one in Excel and many are available online. Once you have randomly assigned participants to groups you need to check whether the intervention and control groups are **well-matched** on characteristics that are important for your population and innovation. If they are not, you can re-randomise the groups until you are satisfied they are well-matched. Remember it is only unfair to conduct a randomised evaluation if you are certain that the intervention is going to be better. If you are sure then you don't need an evaluation.

If randomisation is not possible then you must make sure that the groups are very well-matched. The comparison group should be as similar as possible to the intervention group on prior achievement and relevant demographics (eg, percentage of pupils eligible for free school meals, percentage of pupils with English as an additional language, etc). You should also determine pre-existing **equivalency** on other factors relevant to your innovation. For example, if your innovation is to improve the English oral language skills of EALs, you would need to establish that both groups had similar English oral language at the beginning of the study.

c) Innovation

Describe the innovation clearly so that people who are involved in implementation understand exactly what they are expected to do, and external people who read about the evaluation understand what the innovation involves. Clearly outline:

- What the innovation aims to teach (eg, subject(s), topics, skills, etc.),
- The level at which the intervention is delivered (eg, individuals, classes or whole schools),
- Who will implement the innovation (eg, teachers, TAs, peers, technology),
- Any training staff will receive in delivering the innovation, and any support which will be provided while the innovation is being delivered. Consider what will happen if members of staff leave during the innovation, and how new staff would be trained,
- Any resources available to support the delivery of the innovation, such as manuals or lesson plans (having clear directions from common resources can help to increase consistency with which the innovation is delivered),
- What participants will experience, including:
 - The length of time over which the innovation will be delivered,
 - The number, duration, frequency and timing of innovation delivery,
 - The types of activities participants will engage in,
 - Any expectations beyond the direct delivery of the innovation.

and any other information which is pertinent to your innovation.

You should also outline expectations for control group activity: this may be 'business as usual' (ie, whatever the teacher or school would usually provide), an alternate intervention or a variation of the innovation. Consider how you will avoid contamination of the control group (ie, ensure that they do none of the innovation).

d) Outcome measures

It is important to carry out objective **outcome measures** if you wish to establish the impact of your innovation; people tend to believe that the things they have invested time and effort in have

a positive impact, even when the data shows this is not the case, so relying on participant views is unlikely to give a reliable picture of the impact of your innovation. You should therefore carry out the same measures with your intervention and control groups to compare the progress of pupils who experienced the innovation with those who did not. You need to administer an outcome measure before the innovation starts (a *pre-test*) to establish pupil starting points and how similar the control and intervention groups are. You will also need to administer an outcome measure after the innovation has ended (a *post-test*) to establish how much progress pupils in the intervention and control groups have made.

Both pre- and post-tests should not be too onerous for pupils or teachers; if possible use a common measure that pupils would carry out anyway (eg, typical end-of-year assessments). Measures must be fair to the control group, not assessing content only taught to the intervention group. Where possible, pre- and post-tests should be comparable (eg, a *standardised measure* with equivalent versions). The Education Endowment Foundation's [DIY Evaluation Guide](#) has information about the different types of outcome measure you could use.

You will need to make decisions about:

- Which measures you will use,
- When measures will be administered,
- How measures will be administered,
- How tests will be marked and by whom (if possible, this should be done by someone who does not know the pupils and does not know which condition each pupil is in),
- Whether you will use any non-academic measures (eg, motivation, behaviour, attendance), and how you will maximise the objectivity of any observational measures such as pupil behaviour.

e) Process evaluation

A *process evaluation* provides information about how an innovation was delivered. It is useful to carry out a process evaluation alongside impact evaluations (which consider the impact of the intervention) so that we can understand if the innovation was delivered as intended, obtain teacher and pupil views of the innovation, and participant perceptions of things that went well or could be improved.

To determine if the innovation is delivered as intended you should consider conducting observations of the innovation being implemented. To check that the control group is not receiving elements of the innovation you should conduct observations in the control classes as well. Unobtrusive, structured observations with a checklist or rating scale of critical features of the innovation will be easier to summarise and analyse. Usually observations are conducted after participating staff and pupils have had some experience with the innovation.

To determine staff and pupil perceptions of the innovation, you should consider conducting surveys and/or interviews. The surveys and interviews should include sections with closed questions that are easy to code, making analysis easier, as well as open questions. Surveys and interviews are usually conducted towards the end of the evaluation so that you avoid novelty effects (people liking or disliking something purely because it is new) or learning curve effects (participants not yet fully understanding the innovation).

f) Analyses

Outcomes

Describe how you will analyse your quantitative (numerical) outcome data. You could report the **mean** or **median** scores of the control and intervention groups at pre- and post-tests and calculate **effect sizes** (see below).

If you are going to carry out analyses of impact for any subgroups of pupils (eg, pupils with specific prior attainment, gender, indicators of financial deprivation, etc) decide what analyses you will carry out on which subgroups while planning the evaluation. It is important to decide this in advance, because if you carry out many analyses you are likely to stumble across results that appear to suggest strong impact, but which actually occurred by chance (reporting only the significant results from numerous analyses is known as “cherry picking”).

Process evaluation

Describe how you will analyse data gathered for the process analysis. This may include quantitative analysis of responses to closed questions or data from checklists, and **thematic analysis** of **qualitative data** such as responses to open questions or observational data.

Timeline

Create a detailed timeline or **Gantt chart** for your project plan, including milestones and who is responsible for achieving them. Include this in your agreements with leadership team(s), teachers and, if applying for an innovation evaluation grant, the IEE.

Budget

If you are applying for an innovation evaluation grant, you will need to include a budget in your application. Remember that indirect costs are not allowed and the maximum allowed to be charged to the grant is £5,000 per school for a maximum total of £20,000. Consider how costs which go above this will be covered.

It is also sensible to create a budget during the planning stage if you are not applying for a grant. This will help ensure that you have sufficient funding to complete your evaluation and that you can access funding at appropriate times during the evaluation.

Communication

Decide how and where you will share the findings of your evaluation. This should include writing and publishing an evaluation report. It may also include presenting your findings at meetings or conferences, which could be supported by the Research Schools Network.

Carry out the evaluation

Carry out your evaluation as described in your plan. Remember to complete elements of the process evaluation, such as carrying out lesson observations, as well as administering outcomes measures and delivering the intervention.

Have clear, consistent and frequent communication with everyone involved throughout the evaluation, and address any issues as soon as they arise. You should also closely monitor your budget during the evaluation.

Results

Once you have collected your data, you need to analyse your results. This includes outcome measures and information gathered as part of the process analysis.

a) Outcome measures

Record the data from your outcome measures in a spreadsheet. To maintain confidentiality you should give each pupil an ID number. Keep a record of the students' names and ID numbers securely in a separate location. Each column can be a different outcome that you measured and each pupil (as an anonymous ID number) will have a row.

There is no expectation to conduct fancy high-level statistics within IEE innovation evaluation projects, unless of course this is your cup of tea. It would be good to calculate the mean or median pre- and post-test scores for the control and intervention groups, and to compute effect sizes.

The formula for effect sizes is:

$$\text{Effect Size (ES)} = \frac{(\text{Mean of the intervention group} - \text{Mean of the control group})}{\text{Standard Deviation (SD) of the whole sample}}$$

Steps for computing an Effect Size in Excel

1. Enter outcome data for each pupil into a spreadsheet, with data from intervention and control pupils in different columns. To make it simple and clear, you only need data from the measure for which you are calculating an effect size in this sheet.
2. Compute the mean for the intervention group (in Excel this is called 'AVERAGE' and can be found under 'AutoSum' in 'Formulas').
3. Compute the mean for the control group.
4. Compute the **standard deviation** for the whole sample (in Excel select 'STDEV.S' from 'Statistical' under 'More Options' in 'Formulas', highlight the outcome data for all intervention and control group pupils and click OK).
5. Subtract the average of the control group from the average of the intervention group, and divide that number by the standard deviation of the whole group. That figure is the effect size.

As this is a pilot evaluation, there is a strong likelihood that **tests of statistical significance** will show that the results were due to chance. The smaller the number of participants the more likely that any results will be found to be due to chance, rather than as a result of the intervention. However, the effect size should provide an indication of the potential of the innovation.

You may have decided in the planning stage to examine the results of pre-identified subgroups (eg, by gender, attainment, eligibility for free school meals etc.) to determine if the innovation has a greater impact with certain groups of pupils. You could calculate subgroup mean/medians and effect sizes to compare the scores of subgroup pupils in the intervention group with subgroup pupils in the control group.

Tables, graphs or other infographics are often a concise way to report quantitative results. You can use Excel to create tables or graphs of your findings.

b) Process evaluation analyses

Analyse the data gathered from your process evaluation. When you have numeric data (eg, responses to closed questions or scaling) you can calculate **median** or **mode** of the responses and the effect sizes (see above).

Open-ended responses can usually be coded for easier comparison. For qualitative data that cannot be coded, you can conduct a thematic analysis and triangulate the data (ie, examine open-ended responses to surveys and interviews that point to similar conclusions).

Discussion/Conclusions

a) Interpret your findings

You should consider information from both the impact and process analysis when interpreting the results of your evaluation. Consider whether your findings are similar to or different from the existing evidence base and why this might be the case.

b) Limitations

When drawing conclusions, it is important to consider constraints and methodological limitations of your evaluation which might limit the strength of your findings. There are many factors which could mean that something other than the innovation could have caused the effects you measured (these are called threats to **internal validity**, and you can read more about them [here](#)), such as:

- changes other than the innovation at some or all of the schools,
- lack of similarity the intervention and control groups before the innovation,
- drop-out from one or both of the groups influencing findings,
- the control group receiving elements of the innovation, or
- the control group feeling or being perceived differently because they are not receiving the innovation, leading to changes in control group behaviour or treatment.

You should also consider the settings to which findings can be generalised (the **external validity** of the findings); based on the scale of your study and the characteristics of the schools involved, think about what sort of schools would be likely to find similar results.

When you have decided what conclusions you can draw from your evaluation, you should think about the similarity of your findings to previous research in the area and why you think your findings do or do not support the existing evidence base.

c) Implications for practice

Think about:

- What you will do differently as a result of these findings, and
- Recommendations you would make to other teachers or school leaders on the basis of these findings.

d) Implications for further evaluation

Think about:

- Any further research you would like to carry out as a result of your findings, and
- Recommendations for researchers working in this field.

Communicate your findings

You should share your findings so that others can benefit from your evaluation.

Write a report detailing your evaluation. Below is a sample table of contents for the final report of your evaluation. More information about writing up your evaluation can be found in the document [Writing up your evaluation report](#). For projects that receive an IEE innovation evaluation grant, you will need to submit this report to the IEE within three months of post-testing.

Sample contents for Evaluation Report

Executive summary

- Description of the innovation
- Summary of the evaluation
- Findings
- Costs

Introduction

- Description of the problem
- Review existing evidence
- Propose your innovation (brief description)
- Research questions

Method

- Sample
- Assignment to condition
- Innovation
- Measures
- Process evaluation

- Analyses
 - Outcomes
 - Process analysis
 - Cost

Results

- Outcome findings
- Process evaluation findings

Discussion/Conclusion

- Interpreting your findings
- Limitations
- Implications for practice
- Implications for further evaluation
- Conclusion

References

Appendices

Reports for projects which receive an IEE innovation evaluation grant will be published on the IEE website. It is important to publish your findings, even if results are negative or if you did not complete all elements of the evaluation: sharing negative results will reduce the chance of people trying things which are unlikely to work. If the process evaluation showed that teachers stopped delivering an innovation because they didn't like it, others might consider how it could be improved to increase the chance of delivery.

You could also share your findings by speaking at conferences and events. Your local Research School may be able to support you with this.

Reflections

After you have completed your evaluation, it is useful to reflect on the process. You should consider what went well, what you have learned and what you would do differently in future evaluations (more information can be seen in the document [Writing up your evaluation report](#)). To support reflection and to help the IEE support future projects effectively, leaders of projects granted an innovation evaluation grant will be asked to share reflection under the following headings:

- Carrying out the evaluation
- Budget
- Timeline
- Communication
- Next steps

Ongoing support

For support around planning an evaluation and applying for an innovation evaluation grant contact Alicia Shaw, Research Schools Facilitator at the IEE– alicia.shaw@york.ac.uk 01904 328108. If she can't answer your questions, she will know where you can go for an answer.

Appendix

Example consent letter

Parent/Carer Information Letter and Withdrawal Form

Effective Marking Evaluation

Dear Parent(s)/Carer(s),

This year, The Westwing Academy will be trialling a feedback method called Effective Marking in Year 9 and 10 maths lessons. In Effective Marking, pupils are given structured tasks to identify misunderstandings in maths lessons and are given immediate written feedback to address these misunderstandings. We will be evaluating the impact of Effective Marking on students' learning by comparing the progress of students in maths classes who experience Effective Marking to the progress made by students in classes who do not experience Effective Marking.

This evaluation project is funded by an innovation evaluation grant from the Institute for Effective Education. The Institute for Effective Education is a charity that aims to improve teaching and learning for all children and young people by working closely with schools to support the development and use of effective and efficient approaches.

What would this mean for my child?

All pupils in Years 9 and 10 will participate in maths lessons as usual, and will cover the usual maths curriculum. End-of-unit and end-of-year tests will be carried out as usual, and these assessments will be used as measures for the evaluation. If you **do not** want your child's test scores to be used in the evaluation of Effective Marking, please sign the withdrawal form below.

Confidentiality

We would like to assure you that information from the study will be kept strictly confidential. Students' results will be stored by code number, and any identifiable information will be stored separately from their results. The data will only be used to evaluate the impact of Effective Marking. Anonymised data will be accessed by school staff involved in the evaluation project and may be shared with the Institute for Effective Education to support data analysis. The results of the evaluation will be shared for research purposes, but children will not be identified individually.

Please note, you do not have to agree to your child taking part in the evaluation. If you complete the withdrawal form your child will take part in all science lessons and their science assessments will not be affected. You are free to withdraw your child's data from the evaluation at any time.

We hope that you will agree to your child taking part. If you have any questions about the evaluation, please feel free to contact XXXXX by email (XXXXXXXXXX) or by telephone on XXXXX.

If you **do not** wish your child's data to be included in the evaluation please complete and return the withdrawal form to XXXX as soon as possible/by.....

Thank you for taking the time to read this information.

With best wishes,

Parent/Carer Withdrawal Form
[Effective Marking Evaluation]

I **do not** agree for my child _____'s data to be used in the above evaluation.

(child's name)

Name of Parent: _____

Signature of Parent: _____

Date: _____

Resources

Education Endowment Foundation Guidance Reports

<https://educationendowmentfoundation.org.uk/resources/guidance-reports/>

Education Endowment Foundation. *Teaching and Learning Toolkit*,

<https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit>.

Education Endowment Foundation. (2013). *The DIY Evaluation Guide*,

[https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_DIY_Evaluation_Guide_\(2013\).pdf](https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_DIY_Evaluation_Guide_(2013).pdf)

Institute for Effective Education. *Best Evidence Encyclopaedia*. <http://www.bestevidence.org.uk/>

Institute for Effective Education. *Best Evidence in Brief*. <http://www.beib.org.uk/>

Institute for Effective Education. *Evidence 4 Impact* www.evidence4impact.org.uk

Glossary

Assignment to condition - relates to how participants are allocated to different conditions.

Condition - refers to the group participants are in (eg. they may be in the intervention or control group).

Control group (sometimes called the comparison group) - the participants who do not receive the intervention.

Effect size - a measure of the difference between the outcomes for the participants who experienced the intervention and those who did not (the control group). This is usually between -1.0 and +1.0. If the number is negative, the control group did better than the intervention group. If it is positive, the intervention group did better. The larger the number, the larger the difference between the two groups.

Equivalency - a term used to consider how similar the intervention and control group are before the intervention. As a guide, the the two groups' mean scores should be less than 0.25 standard deviations from each other if they are to be seen as equivalent in relation to that measure. If groups are equivalent, they are sometimes referred to as well matched.

External validity (sometimes called generalisability) - the extent to which research findings can be applied to other situations and populations. More information about threats to external validity and ways to improve external validity can be seen [here](#).

Gantt chart - a visual representation of a timeline which shows the tasks to be completed during certain time periods as horizontal bars.

Impact evaluation - this is the part of the study which considers whether or not the innovation has an impact on pupil achievement or other outcomes.

Implementation fidelity (sometimes referred to as treatment integrity) - refers to whether the innovation was delivered as intended.

Inclusion criteria (sometimes referred to as eligibility criteria) - the characteristics prospective participants must have in order to be included in an evaluation. This may include attainment, demographic characteristics or previous experiences. If your innovation is targeting a specific group of students (eg, those with low prior attainment, poor attendance etc.), the inclusion criteria will include having been identified within this group.

Exclusion criteria - characteristics prospective participants must **not** have if they are to be included in the evaluation. This may include having experienced the innovation in the past or engaging in a similar intervention during the evaluation period as well as demographic, attainment or behavioural characteristics.

Internal validity - the extent to which we can be confident that the innovation has caused any effect identified in the evaluation. There are many threats to the internal validity of an evaluation, and you can read more about them [here](#).

Intervention group (also known as the treatment group or experimental group) - the participants who experience the intervention.

Logic model - a model outlining how your innovation might lead to your hoped-for outcome. Logic models often include the headings “inputs”, “activities”, “outputs”, “outcomes” and “impacts”.

Mean - the average that is computed by adding up the numbers and dividing by the number of numbers. This can be used to find the average of data where the distance between scores is constant (eg, height, speed, test score (depending on the test - this can be controversial!)).

Median - this is the value above which half of the figures lie. This can be used to find the average of data where the distance between scores is not constant (eg, ratings on a scale of 1 - 10, ranking of pupils' performance on a test).

Meta-analysis - methods for quantitatively summarising the effects of a practice or intervention from multiple experiments.

Mixed-method research - studies which use both qualitative and quantitative methods.

Mode - the value that most commonly occurs in a group of scores. This is used as an average for non-numerical, categorical data (eg, participant reports of their favourite element of an intervention).

Outcome measure - a test that can be used to objectively measure performance.

Participant - an individual who takes part in an evaluation.

Pilot study or **pilot evaluation** (sometimes referred to as a feasibility study) - an initial test to assess the feasibility and potential impact of an innovation or theory of action.

Population - all individuals who could have formed part of the sample (ie, all people who meet the inclusion criteria).

Experimentally accessible population - all individuals who could be chosen to participate in the evaluation.

Target population - the total group you wish to understand and generalise your conclusion to.

Post-test - an outcome measure administered after participants have completed the innovation.

Pre-test (sometimes called a baseline measure) - an outcome measure administered before participants begin the innovation.

Process evaluation - the part of the evaluation that collects and analyses data that does not relate to pupil achievement. It assesses factors such as implementation fidelity (whether the innovation was delivered as intended) and attitudes of staff or pupils. It generally consists of conducting observations of the innovation being implemented, observations of the control group, and/or gathering information about the perceptions of staff and pupils participating in the innovation, often by conducting surveys or interviews.

Qualitative data - non-numerical data. This information is usually (but not always) in the form of words (eg, written or transcriptions of spoken answers to open questions, narrative reports of observations).

Qualitative research - studies for which the results are expressed in narrative interpretations of the data.

Quantitative data - data in the form of numbers, which is generally analysed using statistical methods.

Quantitative research - studies which report their findings using numerical data.

Random assignment - a method of assigning research participants to intervention or control groups merely by chance.

Randomised controlled trial (RCT) - a research design in which participants are randomly assigned to participate in an intervention or not. It increases the likelihood that the two groups are equivalent at the beginning so the difference between the groups is not due to factors other than the intervention.

Cluster randomised trial - In this type of study groups of participants, usually schools or classes, are randomly assigned to intervention or control conditions.

Reliability - considers whether the results of an investigation (or an individual measure) are stable and consistent and so would lead to the same results with the same sample on different occasions.

Sample - the group of people who take part in an evaluation. The individuals in the sample are called participants.

Sampling - the process of selecting the sample from the target population.

Standardised measures - outcome measures which are administered and scored in a standardised way, allowing comparisons in performance to be made across large groups of people. These assessments are likely to measure the outcomes for which both the intervention and control groups were preparing their pupils, so they will not favour the intervention participants.

Equivalent versions of standardised measures - some standardised measures have equivalent versions of test papers, which measure the same skills using different questions. These can be administered at different times so different but equivalent tests can be taken as pre- and post-tests.

Standard deviation - this is a measure of the variability of a group of scores. The smaller the number the less variance there is among the scores.

Statistical significance - relates to whether the results are likely to have occurred by chance according to tests of statistical significance.

Systematic review - a review of the research in a particular field. Systematic reviews answer a pre-defined question and are carried out in a systematic way.

Tests of statistical significance - statistical tests which determine whether findings are likely to have been caused by chance, which usually give results as p-values. P-values fall between 0 and 1, and the smaller the number the less likely that the result occurred by chance. If p is less than 0.05 the result is said to be statistically significant and is unlikely to be due to chance.

Thematic analysis - a way of analysing qualitative data which identifies, analyses and records patterns within the data.

Well-matched - see equivalence.