

Innovation evaluations handbook

Institute for Effective Education

February 2017

What do we mean by an innovation evaluation?

The overall goal of the Research Schools Network is to improve the attainment of pupils by increasing the use of evidence-based teaching and learning practices. The goals of the innovation evaluations fund are to increase the evidence base of what works in education by conducting evaluations of innovations of teaching and learning approaches, communicating the findings among the Research Schools Network, and encouraging applications for larger rigorous evaluations of promising approaches.

The innovations that are supported by this funding scheme are 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. Nonetheless, 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.

We have tried to avoid jargon in this handbook and have included a brief glossary in the Appendix to help interpret the research you come across.

Steps in conducting an evaluation

Describe the problem, your chosen innovation, and the background evidence.

Describe the problem or issue that your innovation is designed to address. What challenges do the schools have that need to be addressed? How does existing research evidence suggest that this challenge could be addressed?

Describe the innovation you will evaluate. How will the innovation help improve the problem you have identified? How will this project contribute to the knowledge base of what works in teaching and learning? How will it benefit teachers, learners, other stakeholders?

What existing research evidence supports your view that this innovation will be effective? This might be a review of similar practices. It might be a previous intervention that you are adapting for use with your particular pupils. It should include a logic model, explaining by what mechanisms do you think it works. You might draw on ratings of effective practices from the EEF Teaching and Learning Toolkit (<https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit>), IEE Evidence for Impact (www.evidence4impact.org.uk), 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.

Specific research question or hypothesis

Make a specific prediction about what you expect to happen. What intervention, for how long, on what outcomes, for which pupils? (e.g. Year 6 struggling readers who participate for 6 months in My Favourite Reading Intervention will perform better on KS2 reading comprehension compared to Year 6 struggling readers who participated in the school's regular support for struggling readers (business as usual)).

Method

Sample/Recruiting the participants

What year level(s), gender, percent eligible for free school meals, percent English as a second language, urban/rural, current reading levels, etc. will you recruit? Who will be your comparison group – as similar as possible to the intervention group.

Conduct recruitment meetings to be sure that everyone is clear on what is going to happen. Make sure you have buy-in from the school(s) leadership team.

Have copies of clear agreements signed by all staff and researchers. Have clear, consistent frequent communication. Plan regular meetings for all stakeholders (e.g. leadership team, intervention leaders, teachers, TAs, data collectors, if applicable). Make sure all is clear to reduce the likelihood of participants dropping out of the evaluation.

Consent: - Will you need any consent? If so, opt-in or opt-out? Ensure that the procedures or assessments are not potentially harmful. See appendix A for a sample opt-out letter of consent.

Design

Assignment to condition

Allocating participants randomly to the control or intervention group is preferred but not necessary. 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.

You should assign participants to condition at the level at which the innovation is delivered. For example, if delivered to individuals as in a one-to-one tutoring programme, then you can assign individuals to control and intervention groups. If the innovation is delivered to whole classes then it is better to assign classes to control and intervention groups. If the innovation is a whole-school intervention, then it is best to assign schools.

You can assign intervention and control groups by using a random number generator. There is one in Excel or many are available online.

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 demographics such as percentage of pupils eligible for free school meals or percentage of pupils with English as an additional language.

Outcomes you will measure

Pre-test: If possible use a common measure that is administered by the school to see if your control and intervention groups are equivalent. If there is none, administer a pre-test before allocation to intervention or control or some other way to determine that the pupils are similar to each other at baseline.

Post-tests: Both pre-and post-tests should not be too onerous for pupils or teachers; must be fair to the control group (e.g. typical end of year assessments), not assessing content only taught to the intervention group. Where possible, pre- and post-tests should be comparable (e.g., a standardised test with equivalent versions).

How will the tests be administered? By whom – if possible by someone who does not know the pupils and does not know which condition each pupil is in? Will you also measure non-academic outcomes e.g. motivation, behaviour, attendance? Be wary of online tests unless the schools, teachers, and pupils are used to taking them.

Describe the implementation of the innovation

What the innovation is comprised of (e.g., subject(s), topics, skills, etc.)?

Is the intervention delivered to individuals, classes or whole schools?

Who will implement the innovation – teachers, TA's, technology?

Describe training and follow-up support that staff will participate in (e.g., how long, how often).

Guidance (e.g., manuals, lesson plans) - clear directions to ensure consistency as much as possible.

Additional resources – what else is needed and how will it be obtained, paid for.

What will happen if staff members leave? How will new staff be trained?

Control group activity – describe business as usual or alternate innovation. How will you avoid contamination of control group, that is ensure that they do none of the innovation?

Process evaluation

To determine if intervention was implemented as designed and that the control group was not doing elements the intervention, you should consider conducting observations of the innovation being implemented and 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 the staff and pupils have had some experience with the innovation.

To determine the perceptions of staff and pupils to the innovation, you should consider conducting surveys and/or interviews. The surveys and interviews should also have sections that have closed ended questions that are easy to code, making analysis easier. Surveys and interviews are usually conducted towards the end of the evaluation so that you avoid novelty effects or learning curve effects.

Analyses

You should record the data from your measures in an Excel spreadsheet. To maintain confidentiality you should give each pupil an ID number. Keep a record of the students' names and ID numbers locked away separately.

Each column can be a different outcome that you measured and each pupil (as an anonymous ID number) will have a row.

We do not expect you to conduct fancy high level statistics, unless of course this is your cup of tea. It would be good to compute effect sizes. We'll help you learn how to do that if it's applicable but basically this is the formula:

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

Steps for computing an ES

1. Enter the outcome data for each pupil.
2. Compute an average for the intervention group.
3. Compute an average for control group
4. Subtract the average of the control group from the average of the intervention group and divide that number by the standard deviation. You can get Excel to compute the SD.

If this is a pilot evaluation, there is a strong likelihood that the result is due to chance. The smaller the number of participants the more this is likely. However, the effect size should provide an indication of the potential of the innovation.

Analysing process evaluation: You should have numeric findings where possible

to compute an ES for those as well. Even 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, that is examine open-ended responses to surveys and interviews that point to similar conclusions.

Results

Summarise what your analyses found as clearly as possible. Report both the achievement data and the process evaluation results, where appropriate.

If you examine the results of particular subgroups (e.g., boys and girls, pupils eligible for free school meals), report those here also.

Tables or graphs are often a concise way to report quantitative findings.

Discussion/Conclusion

Interpreting your results: Provide a plain English summary of the findings. explaining whether they are similar to or different from the existing evidence base. You might provide reasons why you think they did or did not support that evidence base.

Limitations: Explain what some of the constraints of your evaluation are. There could be a number of these, which could affect the findings.

1. It might be because the intervention you were testing does not have the impact on the outcomes that you measured.
2. There might be systematic differences between the groups which random allocation or other measures that you used to determine that the groups were as similar as possible did not work. For example, the control group might have better teachers than the intervention group.
3. Many of the pupils or classes dropped out of the study, making the groups unequal.
4. There were differences in the timing of the pre- or post-tests for one group that could have affected the results.

There are many factors that could explain the differences in your results that have nothing to do with the effectiveness of the intervention you are testing.

Implications of your evaluation: What will happen in your school(s) if your innovation is successful? What will happen if it does not improve the outcomes you assess? If successful will you apply for further funding to conduct and larger trial (e.g., to EEF)?

Communication: You will need to submit the final report (see sample table of contents below) to the IEE within 3 months of post-testing. We expect the findings of all projects funded by innovation evaluation grants to be shared on the Research Schools website,

regardless of whether the results are positive or negative. How will else you share your study within the Research Schools network? How will you disseminate your findings more widely?

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 your leadership, teachers, and the IEE.

Budget

Edit the budget in your application and closely monitor it. 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. If costs go above this, how will they be covered?

Ongoing support

Contact Alicia Shaw, the 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.

Writing up your innovation evaluation report

Below is a sample table of contents for the final report of your evaluation. You will need to submit this report to the IEE within 3 months of post-testing.

Executive summary

- Description of the innovation
- Summary of the evaluation
- Effect sizes
- Costs

Introduction

- Detailed description of the innovation
- Review of previous research
- Research Questions

Method

- Sample
- Design

- Measures
- Intervention/procedure
- Analyses

Results

- Achievement findings
- Process evaluation findings

Discussion/Conclusion

- Limitations
- Implications for practice
- Implications for further evaluation

Appendix

- Data that someone can take into a review or that will inform further research

Appendix

Resources

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. *Evidence 4 Impact* www.evidence4impact.org.uk

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/>

Brief glossary

Control Group - the participants who do not engage in the intervention and who are similar to those who do.

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 most often between 0 and +1.0. If the number is negative, the control group did better. If it is positive, the intervention group did better. The larger the number, the larger the difference between the two groups.

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

Mean - the average that is computed by adding up the figures and dividing by the number of figures.

Median - this is the value above which half of the figures lie.

Meta-analysis or systematic review - there are methods for quantitatively summarising the effects of a particular practice or intervention.

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

Mode - The value that most commonly occurs in a group of scores.

.

Process evaluation - This is the part of the study that collects and analyses data that is not pupil achievement. It assesses factors such as implementation fidelity, attitudes of staff or pupils. It generally consists of conducting observations of the innovation being implemented and the control classes and/or assesses the perceptions of staff and pupils participating in the innovation, often by conducting surveys or interviews.

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

Quantitative research - studies for which the results can be determined with numerical data.

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

Randomised controlled trial (RCT) - This is 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 are 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

Standardised measures - these assessments are more 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.

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.

