

Factsheet: How to decide what type of evaluation a programme needs¹

Which **evaluation** design is most appropriate for your programme will depend on the research question you are interested in, what stage your programme is at, what research design options there are, any practical constraints and perhaps what Standard of Evidence you are aiming for. This fact sheet focuses on identifying the preconditions you are facing with your programme (i.e. the research question, the programme stage and the practical constraints), briefly lists the most common research design options² and provides some approximate guidance on which option tends to be appropriate for which precondition (see tables).

Top TSIP evaluation tips

- To identify the best research design for your particular programme, you need to know what question(s) your evaluation needs to answer and what stage of development your programme is at
- Use the tables below – those matching research questions and programme development stage against potential research designs – as a rough guide for your choice in research design for your programme
- There are many practical issues that may affect the choice of your evaluation design, but many of them have a workaround

1. Preconditions

Research question³

It is crucial to decide what question(s) you want to answer with the evaluation in order to identify the most appropriate evaluation design.

Example questions:

- Does the programme work compared to an alternative (e.g. other programme or control group)?
- How does the programme work (e.g. what's the active ingredient)?
- Who benefits most/least from the programme?
- How cost-effective is the programme (e.g. compared to other similar programmes)?
- Which version of the programme works best (e.g. what intervention intensity or 'dosage' is optimal)?

¹ There is a chapter on "Identifying the right evaluation for the policy" on pp. 17-24 in the Magenta book 2011

² For more detailed information on the research design options, please refer to the Research Designs fact sheet.

³ Learn more about research questions in chapter 3, pp. 25-40 in Gorard 2013



- What happened in the programme's implementation (e.g. how many of the recruited beneficiaries attended)?

Programme stage⁴

Your research question(s) will in part depend on which stage your programme is at. The stages below are ordered chronologically, but you may find that your programme jumps back and forth between different stages (e.g. from pilot back to designing) or skips the initial steps because they have already been done. Below is a table outlining which research questions are most common in which programme stage.

1. Designing stage

The programme hasn't been rolled out yet, you are designing your programme and investigating what is already known

2. Feasibility stage

You want to find out if it is feasible to run your programme

3. Pilot stage

You are running your programme for the first time but are still changing several parts of it (i.e. the programme content is not quite final/set in stone yet) and only have a small sample

4. Efficacy stage (focus on internal validity)

Your programme content is mostly final (e.g. has a set protocol – written or unwritten) and is rolled out consistently in 'comfortable' conditions (e.g. in a school you've been working in for years with well-trained staff that you've been working with for years)

5. Effectiveness stage (focus on external validity)

Your programme content is final and is rolled out consistently in more unfamiliar conditions (e.g. in a new school and/or with newly recruited staff) and your results are more likely to be similar to those if someone else (rather than this particular practitioner) is rolling it out somewhere else, so they will be more generalizable.

6. Dissemination stage

Your programme is rolled out on a large scale (by the practitioner and/or other providers, e.g. nationally)

⁴ Learn more about the phases of the research cycle in chapter 2, pp. 14-16 in Gorard 2013

Text and content copyright © 2014 TSIP.

All rights reserved.

Matching example research questions to programme stages

Research Question	Designing stage	Feasibility stage	Pilot stage	Efficacy stage	Effectiveness stage	Dissemination stage
Does it work?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
How does it work?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Who benefits?	<input checked="" type="checkbox"/>	<input type="checkbox"/>				
Is it cost-effective?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
What component is most effective?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
What happens during implementation?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

If you are working towards any of our Standards of Evidence, these will have specific requirements for certain research designs.

2. Design options

Each research design is capable of answering specific research questions and covering the needs of a programme at its current stage. Below is a brief outline of the core design options. Please refer to the fact sheet on research designs for further information.

Type of evaluation

- Process evaluation – “usually in-depth field work conducted as part of a trial to check how well the intervention is implemented (fidelity to treatment) and how participants react to it” (Gorard 2013)
- **Impact** evaluation – usually a quasi-experimental or randomised trial assessing whether a programme works
- Economic evaluation (Cost-benefit or cost-effectiveness analysis) – usually an analysis comparing the financial (and non-financial) benefits against the financial (and non-financial) costs

Research designs (see Research design fact sheet for a more detailed list of options)

The following research designs are most commonly used in social intervention research:

- Literature review
- Observational design
- Quasi-experimental design
- Randomised experiment design

Methodological approach

No approach is universally superior to the other. Depending on your research question, however, you may find that one approach is more appropriate than the other, but sometimes you may find that you need both approaches to answer the question fully.

- **Quantitative⁵**
 - Surveys (quantifiable questionnaires)
 - Administrative/[monitoring](#) data
- **Qualitative⁶**
 - Case studies
 - Interviews
 - Focus groups

Matching example research questions with research design options

Research Question	Process evaluation	Impact evaluation	Economic evaluation	Literature review	Observational design	Quasi-experimental design	Randomised experiment design	Quantitative methods	Qualitative methods
Does it work?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
How does it work?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Who benefits?	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Is it cost-effective?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
What component is most effective?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
What happens during implementation?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

⁵ Learn more about what quantitative methodology includes in our fact sheet on quantitative methods

⁶ Learn more about what qualitative methodology includes in chapter 1, pp. 2-8 in Ritchie 2014

Matching example programme stages with research design options

Programme stage	Process evaluation	Impact evaluation	Economic evaluation	Literature review	Observational design	Quasi-experimental design	Randomised experimental design	Quantitative methods	Qualitative methods
Design	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Feasibility	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Pilot	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Efficacy	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Effectiveness	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Dissemination	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Practical issues⁷

The following practical issues may have a significant **impact** on what is feasible in terms of evaluation design. However, you should keep in mind that for many issues there is a workaround, and so these issues should not immediately be used as a justification not to do more rigorous designs.

- Funding constraints
 - The costliest part of an evaluation is data collection.
 - It is worth considering how much the answer to your research question is worth in terms of financial costs (e.g. good RCT can give better answer than 5 pre-posts, so it may actually be cheaper to do the RCT.)
 - Many quasi-experimental methods actually are similarly (or even more) expensive than an RCT since you have to collect additional data to be able to adjust for **confounding** variables in your analyses since your comparison group is not (statistically) identical to your treatment group at baseline
- Participant recruitment & achieving the needed sample size⁸
 - E.g. For smaller programmes at earlier stages of development, they may simply not be working with enough people to justify attempting a comparison group
- Ethical issues such as informed consent, safeguarding, **equipoise**⁹
 - Interventions may address very sensitive or dangerous issues (e.g. asking perpetrators of domestic abuse about the frequency of assault)
 - It may not be ethical not to give the comparison group any treatment at all (e.g. denying them basic public services), so you will have to change your

⁷ Find more information on practical issues in the decision of choosing an appropriate design on p. 24 in the Magenta book 2011

⁸ Learn more about power analysis and sample size calculations on pp. 375-379 in Boslaugh 2012, or on pp. 127-133 in Togerson 2008

⁹ Learn more about ethical and data protection considerations on pp. 79-80 in the Magenta book 2011



research question from “Does this programme work compared to doing nothing?” to “Is this programme more effective than basic public services?”.

You are then essentially testing for the effectiveness of different intensity levels of interventions (i.e. low-intensity versus high-intensity).

- Politics with stakeholders
 - E.g. some practitioners may already be convinced that their intervention is working and may not see the value in conducting an **impact** evaluation (lack of knowledge)
- Reporting requirements
 - E.g. if the funder wants **outcome indicators** to be measured every 6 months
- Lessons from previous evaluations
 - E.g. if **contamination** was a serious problem in a previous trial, you may want to do a cluster-design this time instead of an individuals-based design (whereby the treatment and comparison groups consist of different clusters (e.g. different schools, prisons etc.) rather than different individuals within the same school or prison)

Books referenced in this fact sheet

Boslaugh, S. (2012), *Statistics in a nutshell*, O'Reilly Sebastopol

Gorard, S. (2013), *Research design: Creating Robust Approaches for the Social Science*, SAGE Publications Ltd. London

HM Treasury (2011), *The Magenta Book – Guidance for Evaluation*, HM Treasury

Ritchie, J., Lewis, J., McNaughton Nicholls, C., Ormston, R. (2014), *Qualitative Research Practice: a guide for social science students & researchers*, SAGE Publications Ltd. London

Togerson, D. J., Togerson, C. J. (2008), *Designing Randomised Trials in Health, Education and the Social Sciences*, Palgrave Macmillan Basingstoke

Glossary

Confounders—A variable associated with cause and outcome; can mask a true relationship between another variable and outcome. E.g. if you are trying to measure the impact of an afterschool maths club on children's numeracy in London, and your comparison group consists of schools in a less affluent area of London, the general quality of the schools' teaching might be a confounding variable – i.e. if numeracy scores improve less among the comparison group, that might be due to a poorer quality of teaching rather than the lack of the afterschool maths club.

Contamination—when some of the comparison group participants directly or indirectly receive the intervention. This tends to lead to underestimating the true effect of the intervention.

Effectiveness—the extent to which a programme/intervention has achieved its objectives under normal conditions in a real-life setting

Equipoise—a state of existing knowledge about an intervention where there is genuine uncertainty over whether the intervention will be beneficial.



Evaluation—the rigorous, scientifically-based collection of information about the activities, characteristics and outcomes of a programme/intervention to determine its merit or worth

External validity—the validity of generalized (causal) inferences in scientific experiments. In other words, it is the extent to which the results of a study can be generalized to other situations and to other people than those in the evaluation.

Impact—the long-term, cumulative effect of programmes/interventions over time on what they ultimately aim to change.

Indicator—a quantitative or qualitative variable that provides a valid and reliable way to measure achievement, assess performance, or reflect changes connected to an intervention.

Inputs—the financial, human, and material resources used in a programme/intervention.

Internal validity—The extent to which differences in outcomes between the intervention and control groups in a clinical study can be confidently attributed to the intervention and not to an alternative explanation, which requires reducing confounding factors and bias to a minimum.

Monitoring—routine tracking and reporting of priority information about a programme / project, its inputs and intended outputs, outcomes and impacts.

Outcome—short-term and medium-term effect of an intervention’s outputs, such as change in knowledge, attitudes, beliefs, behaviours.

Randomised controlled trial¹⁰—a highly rigorous experimental design which is able to attribute program effects exclusively to the programme (i.e. not external factors or ‘noise’) by comparing a treatment group with a (statistically) ‘identical’ comparison group, which is achieved by randomly allocating participants in either treatment or comparison group.

¹⁰ Learn more about the randomised controlled trial in Torgerson

Text and content copyright © 2014 TSIP.

All rights reserved.

An Experimental Method Decision Tree

