Glossary of evaluation terms (alphabetical)

Attrition

Attrition, also known as dropout, occurs when participants fail to complete a post-test or leave a study after they have been assigned to an experimental group. It can lead to a biased estimate of the effect size because those that drop-out are likely to be different from those that stay in. For example, less motivated participants might be more likely to drop out of a treatment. A technique used to address this potential for bias is "intention to treat" analysis, where even those that drop out of the treatment are included in the final analysis.

Bias

A study is biased if its impact estimate varies from the real impact. This variation can be linked to weakness in the implementation or design of the evaluation.

For example, bias can be introduced if participants themselves decide whether to join the treatment or control groups. This ability to "self-select" could mean that, for example, particularly motivated practitioners, or those with more resources and capacity, make their way into the treatment group, while less motivated practitioners, with fewer resources end up in the control group. When this happens, differences in the outcomes of the two groups may be due to these pre-existing features (e.g. more resources or more motivation), not the intervention, and the estimate of the effect size will suffer from bias.

There are many other potential sources of bias, including measurement bias, which is avoided by 'blinding' test delivery and marking, and attrition, which is discussed above.

Blinding

Blinding is where information about the assignment of participants to their experimental group (e.g. control or treatment) is concealed from the evaluator, the participants, or other people involved in the study until it is complete.

Blinding can be introduced at various points in an evaluation. In YEF-funded evaluations the following are usually blinded:

- Randomisation. The person carrying out the randomisation does not know any information that could be used to identify the participants being randomised.

- Analysis. The person carrying out the analysis does not know any information that could be used to find out which participants are in which experimental group.

- Provision of the test. Ideally, the person administering the test does not know whether participants are in the treatment or control group.

- Marking. The marker of the tests does not know whether the test paper belongs to a pupil from the treatment or control group

Failure to blind can introduce bias. For example, a test marker may behave differently if they know that young people are receiving an intervention. If they do not like the intervention, they may subconsciously mark the intervention group lower than the control group. Even if a marker does their best to remain fair and objective, their own preconceptions of an intervention can still affect their marking and introduce bias, without them realising.

Compliance

This is the extent to which the participants complied with the intervention as it was intended. Usually, the intervention deliverer defines compliance for their intervention. This might include attendance at training and quality and quantity of sessions delivered, and might involve thresholds, for example, for optimal and minimal compliance. Compliance analysis estimates the impact of a project on those that complied. This is as opposed to intent to treat analysis that estimates the impact for everyone that received it, including those that did not comply.

Control group

Sometimes called a "comparison group", this group does not receive the intervention being evaluated and allows the evaluator to estimate what would have happened if the treatment group had not received the intervention. The control group should be as similar to the treatment group as possible before the intervention is applied. This can be achieved through random assignment or, if randomisation is not possible, matching (also known as quasi-experimental designs see below). There are several types of control group:

- 'Business-as-usual' control group, which does not receive any additional intervention and continues to operate as usual;

- Waitlist control group, which receives the intervention being evaluated at a later date; and

- Active control group, which receives a different intervention.

Where possible, YEF's evaluators ensure that a permanent control group is in place so that the long-term impact of an intervention can be estimated.

Counterfactual

The outcome for the treatment group if it had not received the intervention is called the counterfactual. If a control group is well constructed, it can be used to estimate the counterfactual.

Efficacy trial

An efficacy trial tests whether an intervention worked under ideal conditions. In practice, YEF efficacy trials aim to test whether an intervention worked under developer-led conditions (with the intervention developer closely involved in delivery). A quantitative impact evaluation is used to assess the impact of the intervention on young people's violence and offending outcomes. An implementation and process evaluation is used to understand how different

aspects of the intervention and its implementation can contribute to successful outcomes. An indicative cost of the intervention is also calculated.

Effectiveness trial

An effectiveness trial tests whether an intervention worked under real-world conditions.

In practice, YEF effectiveness trials aim to test a scalable model of an intervention under everyday conditions (where the developer cannot be closely involved in delivery because of the scale) with a large number of participants and settings, and usually across at least two different geographical regions. A quantitative impact evaluation is used to assess the impact of the intervention on participants' outcomes . An implementation and process evaluation is used to understand how different aspects of the intervention and its implementation can contribute to successful outcomes at scale and in varying contexts. The cost of the intervention at this scale is also calculated.

Effect size

An effect size is an estimate of the size and direction of a change caused by an intervention. It is calculated by dividing the difference between the scores for the intervention group and a control group by the variation in that difference.

Experimental design

A research design where the treatment and control groups are planned to be identical before the intervention is applied. This is usually achieved through random assignment and allows the evaluator to assume that any change in outcomes is due to the intervention, not any pre-existing characteristics.

External validity

Describes the extent to which the results of an evaluation apply to another context. It is also known as 'generalisability'. For example, a study which finds that

an intervention is effective with girls may have poor external validity with boys, because they may respond differently to the intervention materials.

Feasibility study

This is a small scale study to establish the feasibility of an intervention's core activities, as well as its ability to recruit and retain participants. It usually uses a mixed-methods approach and will track service use. More information can be found in Step 4 of EIF's ten steps to evaluation success.[10]

Fidelity

Refers to whether an intervention is being implemented as intended by the developer. If there is low fidelity (teachers, students or schools do not follow the project closely) it is difficult to know whether an intervention is effective or not.

Hawthorne effect

Sometimes called "observer effects", the Hawthorne effect is the phenomenon where participants change their behaviour due to the knowledge that they are being studied. For example, children's behaviour may improve or a practitioner might work harder when an evaluator is observing the session. The presence of Hawthorne effects can lead to biased estimation of the effect size. One way of avoiding the Hawthorne effect is to have an active control group that is also observed

Impact evaluation

A project's impact is the difference between the outcomes achieved by the children who received the intervention and the outcomes of those that did not receive the intervention. Impact evaluation is concerned with identifying the magnitude of this difference (i.e. the effect size) and therefore requires quantitative research.

Implementation and process evaluation (IPE)

IPE is concerned with understanding and explaining why an intervention has or has not been successful, what factors have contributed to its impact and what lessons can be learnt. IPE often uses mixed-methods (including both qualitative and quantitative research) to understand and analyse the views and experiences of key stakeholders (e.g. practitioners, young people and their families).

Intention to treat (ITT) analysis

ITT analysis can reduce the bias introduced by non-compliance and attrition. Analysis is carried out on the groups as they were formed at the point of randomisation. For example, if one of the participants in the intervention group does not comply with the project or the intervention, they are included in the final analysis as if they had received the intervention. ITT provides a credible estimate of how effective the intervention is in a real-world setting. Analysis can also be conducted on those that fully complied with the intervention. This is called 'compliance analysis' and is not commonly used as the primary analysis.

Internal validity

A study has internal validity if the estimate it produces for the difference between the treatment and control group is unbiased.

Intervention

Any project, policy or practice being evaluated.

Literature review

A review of the academic literature on a particular topic.

Logic model

A logic model is a hypothesised description of the chain of causes and effects leading to an outcome of interest. It usually includes inputs, activities (outputs)

and short, medium and long-term outcomes, which are sequenced as they are expected to happen. Short-term outcomes are also sometimes called causal 'mechanisms' that link activities to outcomes.

It is not uncommon for people to use the term 'theory of change' and 'logic model' interchangeably. This is because both explain the theoretical link between an intervention's activities and outcomes. More information can be found in Step 2 of EIF's ten steps to evaluation success.[11]

Matching

A method used to construct a comparison group, matching allows evaluators to control for observable characteristics such as prior offending, age, or family income.

Matching is often used to create a control group when randomisation is not feasible. Participants in the treatment group are matched to others who are not receiving the treatment according to characteristics thought to be relevant to the primary outcome. For example, young people receiving an intervention can be matched with a similar group of young people who have not received it through the Police National Computer that holds information on background characteristics and prior arrests and convictions.

Matching allows the evaluator to assume that any differences in the post test are not due to pre-existing differences in the matched characteristics. For example, if you match young people on prior offending, it is safe to assume that prior offending will not account for differences between the primary outcomes of the different groups. However, matching can only be done on observable characteristics. Some characteristics are unobservable (e.g. attitudes, interaction between family, environment and young person) and are impossible to include in the matching.

Matching can also be used in RCTs to ensure that the groups are balanced. For example, participants can be paired on the basis of prior attainment and then

one from each pair randomly assigned to the treatment group and one to the control group.

Mechanisms

These are the immediate changes or processes that happen as a result of an intervention and that are meant to cause its outcomes. They will describe the way that people are supposed to experience an intervention.

Meta-analysis

A meta-analysis is the systematic analysis of several pre-existing studies of one intervention in order to produce a quantitative estimate of effect size. Meta-analyses also use the techniques of systematic review to decide which studies are included in the analysis. By combining several studies, it is possible to gain a more accurate estimate of an intervention's impact.

Observational study

A study where the assignment of participants to the treatment and comparison groups is not controlled by the evaluator.

Participants

The students, teachers or schools taking part in the trial.

Peer review

This is the evaluation or review of work by one or more people with similar competencies with aim of supporting quality and standards. For example, members of YEF's Panel of Evaluators are asked to peer review YEF evaluation reports written by other panel members.

Pilot

Pilot studies are conducted in a small number of settings to investigate whether an intervention has promise for improving its intended outcomes. It may also involve testing the feasibility of different designs and methods for a future impact evaluation (e.g. methods for recruiting and randomising participants and collecting outcome data). More information can be found in Step 5 of EIF's ten steps to evaluation success.

Post-test or outcome measure

The measure or instrument provided after the intervention, which provides the data used to establish an effect size. These measures should ideally be collected, administered and be marked by someone who is 'blind' to the group allocation.

Power

The power of a study refers to how likely it is to detect an effect size, when there is an effect to be detected. Before starting a study, evaluators estimate the effect size they expect to find. They use this figure to undertake power calculations and estimate the sample size required for an adequately powered study.

Pre-test or baseline measure

A measure or instrument that is carried out before the intervention is introduced.

Primary outcome

The primary outcome is the outcome that determines whether or not an intervention is considered effective, and is the headline finding of an impact evaluation. It should be decided before the trial starts and needs to be stated in the trial registration document. The primary outcome in a YEF-funded evaluation is usually violence or offending. It is good practice to have only one primary outcome.

Qualitative research

Qualitative research is concerned with description. It attempts to explore, describe or explain the social world using language.

Quantitative research

Quantitative research attempts to establish quantities and magnitude. It attempts to explore, describe or explain the social world using numbers.

Quasi-experimental design

An impact evaluation design used when an experimental (e.g. randomised) design is not feasible because the evaluators are not able to control assignment to experimental groups. Quasi-experimental designs use statistical techniques to create treatment and control groups that are as similar as possible before the application of the intervention to the treatment group.

Examples of quasi-experimental designs include matched designs and regression discontinuity designs.

Random assignment

Random assignment is an important feature of randomised controlled trials. It means that the allocation of a participant to the treatment or control groups is due to chance, and not a function of any of their characteristics (either observed or unobserved). If a large enough sample of participants is randomised, the two groups will usually be balanced on every characteristic.

Randomised Controlled Trial (RCT)

The RCT is a type of experimental design where participants are randomly allocated to the treatment and control groups. Random assignment allows the evaluator to assume that there are no observable or unobservable differences between the two groups that could affect the primary outcome, and any effect size is due to the intervention received by the treatment group.

Random assignment is used to deal with the problem of selection bias, which occurs when the way in which participants are assigned to experimental groups biases the findings of the study. For example, if an evaluator allows participants to volunteer for the treatment group and fills the control group from the pool of participants that did not volunteer, any difference in the primary outcome could be due to pre-existing characteristics and motivation of the participants that volunteered. Participants that volunteered for the treatment group are likely to be more motivated and engaged, and these features could explain any improvements observed, not the intervention.

Reporting

YEF reports are based on best practice guidelines for transparent reporting of evaluations and impact evaluation reports will include attrition rates and other sources of bias.[13] Impact evaluation results will be reported on an 'intent to treat' basis, where all outcomes, including those from participants who dropped out are included. The effect size and uncertainty around that effect will be reported alongside commentary about how this fits with the existing evidence.

Sample size

The number of participants in the study.

Systematic review

A synthesis of the research evidence on a particular topic, which uses strict criteria to exclude studies that do not fit certain methodological requirements. Systematic reviews that provide a quantitative estimate of an effect size are called meta-analyses.

Theory of change

A theory of change explains the rationale for why an intervention is needed and explains the theory for how it links to its intended outcomes. A good theory of change will identify short and long-term goals that are important and justify this with links to scientific literature.

It is not uncommon for people to use the term 'theory of change' and 'logic model' interchangeably. This is because both explain the theoretical link between an intervention's activities and outcomes. However, while a logic model is primarily concerned with how an intervention will achieve its outcomes, a theory of change is also concerned with why this is important. More information can be found in Step 1 of EIF's ten steps to evaluation success.[1]

Treatment group

The group of pupils, classes or schools that receive the intervention.

[1] https://www.eif.org.uk/resource/10-steps-for-evaluation-success