

## Evaluation Design Presenters Notes

### Slide 1:

- **What do you want to know?**
- **To improve or to prove?**
- **Interested in processes or outcomes?**

The first question to ask when thinking about evaluation design is, 'what do I want to know?'  
**What is the purpose of your evaluation?**

#### **Is it to improve or to prove?**

If you are looking for information to help you improve an intervention then you will need a formative evaluation design. If you are looking for information to help you to prove your intervention was a success, then you will need a summative design.

Ask the group if they can tell you anything about what is meant by formative and summative?

Explain that formative evaluation is conducted in the early stages of an intervention, and results are fed back whilst the intervention is still being delivered – with the aim of making real-time changes.

In a summative evaluation, results do not get fed back until the end – until after the intervention has finished or is well bedded in. In a summative evaluation you are not primarily looking to change the intervention. Of course results may lead to changes being made the next time the intervention is delivered.

#### **Are you interested in process or outcome data?**

Again, ask group if they can tell you the difference between process and outcome data.

Emphasise that you can collect data at different stages of the intervention.

Explain that process data looks at the delivery mechanisms of the intervention – did everything happen as it should have done? Process evaluation focuses on the input and output stages. It does not measure outcomes. For example: did all the resources delivered to third party agents, get handed out as you had envisioned?

Outcome evaluation looks at the post-intervention stage – the effects of your outputs. Outcomes can be measured immediately after an intervention has been delivered (short-term such as changes in knowledge and understanding), or require long-term follow up (such as behaviour change measured by use of mobile phones while driving, or number of traffic violations involved in).

A formative evaluation can include short-term outcomes, as well as collecting process data. A summative evaluation looks at longer-term outcomes, but can also include process data.

Ideally you should include both process and outcome data in any evaluation you do.

## Slide 2: Key Design Terms

- **External Validity – can you generalise results?**
- **Internal Validity – alternative explanations?**
- **Reliability – could study be replicated?**
- **Bias – researcher and participant**

These are terms that you will come across in looking critically at evaluations. These factors all affect the strength of an evaluation's findings and how confident you can be about those findings. You need to make your evaluation results as defensible as possible.

**External validity** – only if you have used a random sample can you claim generalisability, i.e. that your findings can be generalised to the whole population – that the whole population would have responded the same as those in your sample. In a non-random sample there could be something very different about the characteristics of those taking part in the intervention, and those who do not.

**Internal validity** – was it your intervention that caused the change you found, or something else? There could be other reasons why the situation you wanted to change improved or worsened. These threats to internal validity (also known as confounding variables), can be relatively simple like maturation, i.e. that as participants age, their views naturally change over time. This is particularly important therefore in evaluation's involving young people such as young drivers – if you are doing a long term follow-up study, can you be sure that it was the intervention that caused the observed change, or simply the effects of growing older? Maturation is similar to regression in that if people hold extreme views at one point in time, it is possible that those views will be less extreme when asked again – i.e. regress towards the average strength of view.

Other threats include:

- History/External Context – did something happen to change participants' views in-between the intervention and your outcome measures? For instance, if you're surveying attitudes on speeding enforcement, a serious local or high-profile crash or court case involving a speeding driver may affect your responses.
- Individual differences, i.e. selection bias. Was there something particular about the people who took part in your intervention that meant there was a positive result? For example, were the riders who took up the free post-test motorcycle training, those who are the more cautious riders anyway, or those with particularly safety conscious families? Or were the people in your sample of a similar certain age or education background?

**Reliability** - If somebody else conducted the evaluation again, under the same type of conditions, they should get the same results. If the results can be replicated by others, then the findings are strengthened.

**Bias** – Studies should report any biases that may have affected results. Bias will always be present – for example, the Hawthorne effect where participants will alter their behaviour simply because they know they are part of a study, and are a 'special' group – they may try extra hard. Equally, members of a control group may try extra hard to show that they are no worse off than their counterparts receiving the intervention.

As mentioned above, selection bias is where participants have not been randomly selected, so they may differ from the general population, and/or a comparison group.

Another common form of bias is social response bias – where participants want to please the researcher, and give what they think are the socially desirable answers to questions, instead of responding honestly.

Researcher bias may be prevalent when the person conducting the study has an expectation about, or a vested interest in, what the results of the evaluation will be. It is always preferable for the evaluator to be independent and objective.

### Slide 3: Common Outcome Designs

- **Randomised Controlled Trial**
- **Quasi-experiment**
- **Pre- and Post-**
- **Post-only**

The purpose of this slide is to go through the main outcome designs – emphasise that they are outcome designs, looking at post-intervention effects.

Describe the designs and go through some strengths and weaknesses of each, asking the delegates if they can explain the designs before you do.

**RCT** – The evaluator randomly assigns participants, who meet a set selection criteria, to either an intervention or a control group, before the intervention begins. Experimental designs can be pre and post or post only. The advantage of pre and post, over post only, is that although participants are randomly recruited to either intervention or control group, with a post only design you cannot be sure that both groups were similar to begin with. Even when randomised it could be the case that by pure chance, the participants in the intervention group have less tolerance of drink-driving, for instance.

RCT is a robust method but it is unusual to have the ability (with ethical approval), in social interventions, to randomise participants to intervention or control groups. It is also an expensive and time consuming design.

**Quasi-experiment** – The same as an RCT except there is no randomisation. Instead of a control group, you use a ‘comparison group’ – i.e. a similar group of participants who may be matched by age, sex, locality etc.

The comparison group helps to determine the amount of change that can be attributable to the intervention – i.e. you can see what change would have occurred without the intervention and simply deduct that from the change seen in the intervention group.

You need to be able to re-contact the participants as you are taking measurements at two different points in time. With a comparison group pre and post design, this means taking a total of four measurements, at two time points. Thus time and labour intensive.

**Pre- and Post-** - You take measurements from one group only – once before, and once after the intervention. With the pre (baseline) measurement, this design allows you to know whether or not there was change, but without a control or comparison group, it is not possible to say that the change was caused by the intervention, and nothing else. Fairly cheap design to implement though as only one group of participants involved.

**Post-only** – You take one measurement from one group, only *after* the intervention has been completed. A very cheap, quick and easy design but as you have no baseline (pre) data, you do not know if there has been any change in what you are measuring. You cannot be sure if the participants’ attitude/knowledge/behaviour etc. that you are measuring after the intervention, is any different to how it was before.

This design is only really useful if you have other data to give you some idea of a baseline measurement, e.g. records from a previous course, or, if you have good reason to believe that the participants' attitude/knowledge/behaviour etc. is at a low starting point.

#### **Slide 4: Design Activity**

- **Look at the 5 designs on your sheet. Rate them for strength: give the most robust design a 5, and the least robust, a 1.**
- **Explain your ratings and give some disadvantages of each design.**
- **Identify whether each design is an experiment, a quasi-experiment, or a non-experiment**

Tell the delegates, in groups, to look at the design activity in their packs. Give instructions as above – rate the designs for robustness from 1 (weak) to 5 (strong).

Tell them that they should be able to explain their choices, and also be able to give at least one disadvantage of each design.

Also ask them to be able to say if the designs are an experiment (Design 3: the RCT), a quasi-experiment (Design 5: pre and post test with a comparison group), or a non-experiment (all the others: 1, 2, and 4).

Feedback as a whole group by asking them to shout out their ratings, explanations, and disadvantages. Check for any differences of opinion and correct any misunderstandings.

Re-iterate that the best design will always depend upon their own individual circumstances and resources.