

Evaluation - natural language questions

Paul Ekblom

V1.1 13-07-21

1 Introduction

At the EU Crime Prevention Network task force meeting of 21 June 2021 it was agreed that I set out a series of key questions that evaluators should ask before, during or after a preventive project. The questions should be posed in *natural language* terms so they are as accessible as possible to evaluators across the entire range of experience and sophistication, ranging from the self-evaluating ground-level practitioner to the advanced professional researcher.

What differs between these levels is not the questions, but the way in which they are answered – for example, different levels on the Maryland or similar scales of methodological rigour. However, the questions should be carefully drafted and supported by a glossary that is internally consistent. This ‘definition in depth’ will help in both understanding, training and translation of results between EU languages (perhaps, eventually, Artificial Intelligence-based translation and search – see e.g. [public health equivalent](#)). Some of the questions will be relevant to all instances of evaluation; others will be more appropriate to particular circumstances, for example depending on whether the evaluation results are to be used for purely local accountability and decision-making (continue, expand, modify or abandon the project?), or contribute to a national or international programme of prevention, in which issues of infrastructure, replication in different contexts etc are to be considered.

Evaluation covers both impact and process. In practice, there is overlap in the set of questions to be asked in each type – particularly so with Realist evaluations. Whatever the case, this is all the more reason to endeavour to conduct both impact and process evaluation together.

Most of what follows is developed from the 5Is book, especially Chapter 15.

I have tended to use ‘project’ generically. Most of the questions will also be relevant, at a more strategic level, to **programme** evaluation, although there may be additions/deletions/modifications needed. We could develop separate lists for service, project, programme and policy evaluation.

There are a lot of questions here! It may very well be that we decide to reduce the number that we inflict upon evaluators, especially practitioner-evaluators. Alternatively, we could filter the questions so that we had a minimum set for novices/non-professional evaluators, a larger set for people of intermediate skill/resources and a maximum set (‘the works’) for professional/academic evaluators.

With some items, I supply a brief explanation or expansion of the main question, or examples. This could be taken further, e.g. fuller explanation of what it means and why it’s necessary, how it relates to practice, programme or policy levels, guidance on ways to answer the question, using different levels of rigour according to purpose of evaluation...

2 Impact evaluation

Impact evaluation looks at the relationship between **inputs** of effort, resources etc, **outputs** – the final link in the sequence of actions undertaken by the project team that are intended to have a beneficial crime prevention effect in the real world – and **outcomes** – the changes actually achieved that are attributable to the outputs and perhaps to wider influences of the project action (e.g. publicity, stigmatisation).

Intermediate outcomes are those which result from the outputs but are only part-way towards achieving the **ultimate outcomes**, the desired changes – for example, the number of cyclists that lock their bikes securely following installation of a new design of bicycle stand is an intermediate outcome; the fall in the number of bikes stolen is the ultimate outcome (or if the top-level goal is actually increased use of bikes, an increase in ridership). Intermediate outcome indicators are useful in giving early warning of success (or failure), confirming (or refuting) the logic model assumed for the project, and substituting for ultimate outcome indicators (for

example where crime events measured are very rare, as with terrorist attacks; or ultimate outcome indicators are slow and costly to measure).

2.1 Aims of the crime prevention project

It is necessary to collect information on the aims of the project being evaluated, and how the intervention/s are intended to have their effect, before any evaluation can be planned, designed, executed, analysed or interpreted. Being able to set out this information in advance of planning the project and the evaluation, not only makes for a better evaluation, but helps to ensure that the project itself will be worth evaluating.

2.1.1 What are the **aims/objectives** of the **intervention** – the changes in the real world intended to result from the project?

2.1.2 Are the outputs, and the outcomes, **universal, selective or indicated**?

Outputs and outcomes may differ here. It may be, say, that the outputs are indicated (such as treating persistent offenders) but the outcomes may have effects on the crimes suffered by the whole population.

2.1.3 How can these changes be expressed as measurable **outcomes**?

2.1.4 How do these aims follow, in terms of the rationale of **cause-and-effect**, from the outputs of the intervention?

2.1.4.1 In particular, is the intervention based on any particular **theoretical principles**? And are there particular **contextual factors** that – it is anticipated – need to be present (or absent) for those intervention principles to successfully have their effect on crime?

There are many causal factors that come together to make criminal events happen, whether they are present in the immediate environment or otherwise endow individuals or groups with criminal predisposition, motivation, intent and capability. Crime prevention interventions try to modify some of these factors, in the environment or relating to the offender – blocking, weakening, changing or deflecting them so that criminal events are less likely to happen, or to be so harmful. The intervention principles or mechanisms – how the prevention intervention has its effect on particular causes of crime – often depend on the context of other causes that may be present in the background. For example, neighbourhood watch may only work as an intervention, in a context where social cohesion is already strong.

2.2 Context of the evaluation itself

2.2.1 Is the evaluation **internal or external** to the implementing organisation?

2.2.2 Is the evaluation **independent** or not, e.g. it is funded from a separate source to the project itself?

2.2.3 What **type/s of evaluators** are involved – academic, commercial consultant, practitioner?

2.2.4 Is the evaluation **formative** (continuous feeding back of emerging findings to the implementation team as the project unfolds) or **summative** (the evaluators make minimal contact with the project team until the project is finished/the findings are revealed)?

Constant monitoring or evaluation at set points during the project's progress helps in identifying problems and intervening before the project is irreversibly affected. Hence continuous evaluation strengthens the reliability of the project's final results; and may also ensure, by adjusting implementation and involvement actions, that the project actually achieves an impact that is worth evaluating in the first place.

- 2.2.5 Is the evaluation being done as a matter of managerial **routine**, or is this a special **one-off** exercise?
- 2.2.6 Is the evaluation intended purely for **local** use (e.g. guiding the decision to continue, expand, modify or terminate the project) or for **wider** purposes (e.g. feeding back on programme performance, contributing to collective knowledge of what works)?
- 2.2.7 Is the evaluation **prospective** (planned and started in advance of the project) or **retrospective** (planned and executed wholly after the project has started, or maybe even finished)?
- 2.2.8 Does the evaluation cover **impact, process or both**?
- 2.2.9 Were there any issues of achieving a **climate of understanding and acceptance** of the proposal to undertake an impact evaluation with stakeholders – in advance at planning stage, maintaining it during execution, and in retrospect when presenting results?

2.3 Technical questions on the evaluation methodology

- 2.3.1 What is the overall **approach** of the evaluation - e.g. Realistic, Theories of Change, Experimental, Qualitative?
- 2.3.2 What is the evaluation **design**, e.g. simple before-after comparison, case-control, QE, RCT? How does this relate to methodological **quality scales** e.g. Maryland scale?
- 2.3.3 What are the basic **parameters** of the evaluation design such as output measures, and intermediate and ultimate outcome indicators (e.g. self-reported offending, police recorded crime figures), reference or target population, sample size and units (e.g. individuals, families, neighbourhoods), time periods (e.g. before, during, after, follow-up)?
- 2.3.4 What is the **timescale** of the evaluation?

This can cover the timing of planning of the preventive action, implementation, time for the intervention to have an effect, time to measure changes in outcome indicators, and time to analyse and deliver findings.

- 2.3.5 What methods of **statistical testing** are used and why?

This refers to issues such as power considerations, bias, etc.

- 2.3.6 What **problems, issues and tradeoffs** were encountered in setting/applying these designs, parameters and tests? Were there any **practical resolutions** worth sharing?

2.4 Overview of the action that actually took place

This information, and much more, would usually be gathered for a **process evaluation**, but what follows is the minimum necessary for enabling useful conclusions to be drawn from an impact evaluation.

- 2.4.1 What were the **characteristics** of the problem, places and/or people being **targeted** by the programme?
- 2.4.2 What were the **outputs achieved** by the intervention?

By outputs is meant the final actions or products generated by the project, e.g. number of homes secured, number of young people attending remedial literacy class.

2.4.2.1 Were the planned **output objectives met** in terms of quality and quantity?

For example, did the intervention meet the fidelity standards of the relevant programme (e.g. quality of security fittings and installation; quality of treatment protocol for offenders)? Are the intended recipients of a programme accessing the services being provided, do they remain in contact with the programme and does the programme meet the needs of participants?

2.4.3 Were the **contextual factors or background conditions**, that were originally judged necessary for the intervention to work, actually **in place** during the time of the project?

These can be listed/described under Intelligence, Intervention, Implementation and Involvement.

2.4.4 What were the **ingredients** that were necessary or sufficient to enable these achievements, in terms of practical **implementation** activities, and **involvement** of other actors (individuals, groups, communities, organisations or networks) who were supposed to deliver or help support the action?

These can be listed/described under Intelligence, Intervention, Implementation and Involvement.

2.5 Key impact evaluation questions

2.5.1 Was there a statistically significant **change** in **outcome indicators** relating to crime, safety and other benefits, over the timescales of interest?

We can distinguish between **intermediate** and **ultimate** outcome indicators here; and between short- and long-term effects. Intermediate outcome indicators could include for example changes in participants' knowledge, attitudes, skills or behaviour; ultimate outcome indicators should ideally range beyond crime statistics and perceptions of safety to include wider consequential benefits.

2.5.2 If so, how much of this change can be **attributed** to the outputs of the preventive intervention – as opposed to being due to background trends, coincidental events and 'masking or mimicking' effects such as increased reporting of crimes, regression to the mean and maturation?

Depending on the sophistication of the evaluation, answering this can involve simply relying on a 'control' design, presenting a few elementary cross-checks or a thorough exploration of alternative explanations; quantitatively speaking, answers can range from percent changes to standardised effect sizes.

- 2.5.2.1 To what extent did the project meet its stated **outcome objectives**?
- 2.5.2.2 With multiple outcome objectives, was there any **weighting** to accord different values or levels of priority on each?
- 2.5.3 With multiple sites and/or individuals studied, were the effects confined to **subsets**, e.g. only to highly-cohesive areas, or only to offenders with supportive families?
- 2.5.4 Were there any **adaptive reactions to the intervention** e.g. by offenders showing diffusion of benefit, displacement, longer-term evolution of countermoves and offender replacement (i.e. arrest Mr Big the drug dealer and Mr Notsobig takes his place)? Did other parties adapt such as potential **victims** showing 'conservation of risk' (for example relaxing their guard on where to park, in the belief that their immobiliser would protect their car from theft)?
- 2.5.5 How did the intervention **work** – by what **theoretical principles or mechanisms**, and dependent on what **contextual contributions**? Were the latter **essential** to achieving any impact, or did they merely **boost** the impact?
 - 2.5.5.1 Did the contextual conditions shape (moderate) the nature of the outcome?
 - 2.5.5.2 Were there any **adverse** contextual conditions that reduced the impact?
- 2.5.6 With multiple interventions, **which ingredients/components** were essential to any impact? Which just boosted impact?
- 2.5.7 Were there any **harmful side effects** – on crime and safety (e.g. the intervention made another crime type easier, widened the net for involvement of young people in criminal justice system or increased fear or inconvenience); and beyond (e.g. conflicts with other policy aims such as sustainability or inclusion)?
- 2.5.8 Were there any **beneficial side effects**, e.g. to project staff, other stakeholders or government? Did these come from the Intervention method itself or from Implementation and Involvement actions? In improvement terms, did the trial suggest how harmful side-effects could be reduced or avoided and benefits increased?

Assembling a cumulative list of possible harms and benefits, and how to mitigate the former and boost the latter, can facilitate designs of future interventions and future evaluations.

- 2.5.9 How big and how cost-effective was the **gross attributable change** in the outcome indicators? And how big was the **net attributable change** taking account of offender adaptations and other side-effects on crime?

In the course of estimating cost-effectiveness, other useful economic indicators can be produced, e.g. economy, efficiency.

- 2.5.10 How did these changes translate into **benefits**?

At the very least this can list the public and private cost-savings on crimes prevented; perhaps also knock-on benefits of education or area regeneration.

- 2.5.11 How durable or sustainable was the impact?

This partly depends on the length of the follow-up period of the evaluation. But investigation of mechanisms can indicate the likely future durability of Intervention (e.g. CCTV that works by arrest and conviction may have a longer lasting effect than if it worked merely by deterrence through heightening perceived risk), and

sustainability of Implementation (e.g. how long could payment of extra police overtime be maintained?) or Involvement (e.g. would neighbourhood watch members lose interest if crimes became rare?).

2.5.12 How far did the intervention meet its **aims** and any **quantitative targets**?

2.5.13 What were the **limitations on performance** and how might these be alleviated through **improvements** made to the design, implementation and management of the project?

2.6 Wider performance/selection indicators

A major purpose of evaluation is to share knowledge of crime prevention performance, principally to guide the choice by other practitioners, programme managers or policymakers as to what interventions to select for replication or roll-out. The basic impact evaluation questions (above) supply the core of the information needed to make these choices, but other information is important.

2.6.1 How **responsive** and **scalable** to crime/safety problems was the preventive action?

2.6.1.1 How far did the preventive action address key **priorities** of security/safety in terms of the severity of **consequences** of crime/safety problems?

2.6.1.2 How accurately/efficiently did the preventive action **target** the **needs** of victims and of wider society?

Targeting can be done in various ways: on units at different ecological levels (individuals, groups, organisations, communities, networks, places, products); intervening universally, selectively or on indicated people/places as appropriate; on geographical levels from local to international.

2.6.1.3 What is the extent of **coverage** on the ground – what **proportion** of a given crime problem does the preventive intervention aim to tackle?

There are trade-offs here, for example between focusing on very high-crime areas can be cost-effective, but giving wider coverage of lower-crime areas can tackle a greater proportion of the crime problem regionally or nationally.

2.6.1.4 What is the **scope** of the preventive action, in terms of the range of different crime problems tackled – from e.g. juvenile crime, to theft of luggage at Heathrow Airport.

2.6.2 Over what **timescale** did the Implementation occur, did the Intervention take effect, did the Impact reliably become apparent?

2.6.3 How **legitimate or acceptable** were the preventive actions, within the wider population, within minority subgroups, or even among offenders?

3 Process evaluation

Process evaluation can be defined as the action of evaluating a particular process, against criteria of achievement and/or quality. In turn, process is an organised sequence of actions which together take inputs of various material, human and informational resources and lead via intermediate stages and transformations to desired outputs. In crime prevention, process evaluation covers all the tasks of the 'preventive process'. The process is set out in a simple but very limited way in the SARA model of Problem-Oriented Policing (Scanning, Analysis, Response, Assessment). However, this is insufficiently detailed to systematically capture and synthesise the practical knowledge and know-how to be gained from process evaluation, and to organise it for efficient retrieval on-the-job, and use in training. An attempt to design a process model to serve this purpose is the 5Is framework (Intelligence, Intervention, Implementation, Involvement, Impact and process evaluation – and many detailed sub-headings of these tasks).

As already said, some of the questions that follow overlap with those of impact evaluation – which requires a certain minimum of process information – but here the detail is greater.

3.1 For each main task (Intelligence, Intervention, Implementation, Involvement), or their subsidiary tasks:

3.1.1 What was the **success or failure of implementing each task?**

E.g. What was the nature and extent of stakeholder **Involvement** in all stages/aspects of the programme? In more detail, under Involvement, how well did **partnership** processes work?

3.1.1.1 Were any **output targets or **quality standards** (fidelity criteria) met or missed?**

3.1.1.2 Were any existing good/best practice **benchmarks applied during the design of the Intervention and its Implementation, and/or new ones indicated?**

Deviations from good/best practice may or may not have been planned/justified in advance, or deliberately or unintentionally put in place during implementation. Whether this resulted in success or failure of the project, learning opportunities can be exploited here.

3.1.2 Whether or not the task was successfully implemented to an adequate level of quality, did it deliver the desired immediate **result?**

This could be an **internal** result (like successfully mobilising a young person to join the youth centre), or an **output** to the external world (such as the young person leaving the centre with employment-facilitating skills; in impact evaluation terms this would also count as an intermediate outcome).

3.1.3 Did the task, in achieving or failing in its immediate objectives, engender positive or negative **side-effects?**

Side-effects could be in **crime** (for example, simply targeting a youth for intervention could give them a 'badge of honour' among peers, confirming criminal identity); or in **other spheres** (CCTV could, depending on context, attract or deter commercial tenants in a shopping centre).

- 3.1.4 How could the task be made to work **better** (or even work at all)?
- 3.1.4.1 What were the **enablers and constraints, conflicts, trade-offs and synergies, collaboration and competition** behind the successes and failures of implementation and operation of the project, that would need addressing for improvements to be possible?
- 3.1.4.2 What was the **cost** associated with the operation of the project? Was the project adequately resourced?
- 3.1.4.3 At a more strategic level, how appropriate were the **governance** arrangements, **operating guidelines** and where applicable, **legislative framework** in supporting the operation of a programme?
- 3.1.4.4 How did these issues relate to the **context** of the action described?
- 3.1.4.5 If the task was **replicated** elsewhere, what **contextual conditions** would be conducive to handling the issues successfully? Or likely to lead to failure?
- 3.1.5 How were the process problems encountered **resolved** or **avoided**; and how were **benefits** capitalised on, **side-effects** mitigated, and **failures** coped with?
- 3.1.6 Were there any generic **qualities of implementation and involvement** demonstrated in the project, such as **adaptability** and **improvement, responsiveness and deliverability**; and how might these be replicated?
- 3.1.7 Were there any **innovative** aspects of the tasks undertaken, and how well did they perform?
- 3.1.7.1 Was there anything worth reporting about the **innovation process** itself?
For example, the use of co-design techniques, design student classes.
- 3.1.7.2 Were there any circumstances which **helped** innovation, or **hindered** it?
For example, the risk appetite of project management.

4 Questions common to impact and process evaluation

Some of the following questions relate to straightforward **documentation** of how the evaluation was done, to provide evidence e.g. of the rigour and scope of the evaluation. Others relate to how the experience of the evaluation exercise might lead to improvements and innovations in **evaluation technique**.

- 4.1 What were the overall **methods** of evaluation/quality assurance themselves (e.g. observation, interview, document analysis)?
 - 4.1.1 How **rigorously conducted** were these methods, and what were any possible **biases** and **limitations** in the results?
- 4.2 How were the **evaluation problems** encountered resolved or avoided, and how were **benefits** capitalised on, and **failures** coped with?
- 4.3 Were any of the evaluation techniques **innovative**?
- 4.4 How well did the evaluation techniques **work**, how could they be **improved**, and what were the **limiting conditions** (e.g. only suitable for particular contexts)?
- 4.5 Were the evaluation techniques, and their realisation in practice, **appropriate for the purposes and context** of this evaluation exercise (see 2.2 above)? Or were they, for example, too costly or slow?
- 4.6 Did the mere **presence** of the evaluation, or the way it was **arranged and conducted**, make any difference to the selection, planning, siting or performance of the **project** itself?

4.7 General items on failure

An evaluation can sometimes show **negative or inconclusive results**. To the extent that the project, and the evaluation, have been well-planned and well-implemented, this is less likely, but still remains a possibility. The delicate nature of many crime prevention mechanisms, and their dependence on a range of social or other contextual factors, means **failure is always a possibility**. Particularly where **innovation** is concerned, failure is a risk that has to be accepted, otherwise prevention will never be able to make advances to keep up with changing circumstances and adaptive offenders.

Failures, whilst unwelcome, can be very **instructive** if properly investigated. They can reveal, for example, what **not to repeat** in future, what **ingredients for success** were missing, what **theories** may be revealed as invalid and even whether, in the midst of failure, certain practices developed in a project may be worth **salvaging**. For example, a burglary project that failed to cut crime may nevertheless have developed a really useful way to mobilise residents, that is worth retaining and sharing.

Information on the **fact** of failure comes best from impact evaluation; **why** the failure happened, and how to **avoid** it in future, from process evaluation.

- 4.7.1 If the outputs and/or outcomes were not achieved, or achieved in only a limited way, what were the **causes of failure**?
 - 4.7.1.1 Were there failures of **Intelligence, Intervention, Implementation or Involvement** -and if so, of which subsidiary tasks?
 - 4.7.1.2 Was there a **failure of the evaluation itself** to measure the outputs and/or outcomes properly, and make suitable inferences?

For example, there may have been a lack of statistical power to detect changes in crime of a meaningful size.

4.7.2 Are any of these failures susceptible to **improvement** in subsequent projects/ evaluations?

4.7.3 Were there **salvageable elements** of the action that did seem to succeed, for example in mobilisation of people to implement preventive action?

It helps to identify such elements to preserve (or avoid) if the description of the project action is done in a way that is structured and modular. x

4.8 Evaluation of supporting action – capacity-building, infrastructure, contextual improvements, delivery packages and process models

Most crime preventive action is *operational* – that is, intended to have a direct and measurable impact on ultimate outcomes, such as levels of crime or fear, concerning a specific problem in a specific locality. Sometimes, however, a project may be intended to have an *indirect* effect. Such actions include:

- *Capacity-building* – improving *operational capacity* through knowledge management (involving training, knowledge bases and/or toolkits), data provision and other analytic infrastructure within professional services such as the police or community crime prevention groups.
- Building *innovative capacity*.
- *Motivational approaches* such as the EU Crime Prevention Awards or the Goldstein or Tilley Awards for Problem-Oriented policing.
- Improving the *contextual conditions* for helping crime prevention interventions succeed – perhaps by boosting social cohesion.
- Establishing a programme to roll out a *replicative delivery package* which sets up or boosts the capacity of local teams to generate and undertake good quality preventive projects or services themselves. One example is the [Communities That Care](#) programme which aims to generate rigorous evidence-based action improving the lives and behavioural outcomes for young people, customised to local problems and context. Another is the work of the [Center for Problem-Oriented Policing](#) which aims to establish, embed and support POP practice among police forces internationally.
- Developing *process models* such as SARA and 5Is which aim to support and extend performance of practitioners.

Such capacity- and infrastructure-building efforts, programmatic packages and generative process models can themselves be evaluated (e.g. the 2020 [systematic review of Problem-Oriented Policing](#)). The kinds of evaluation questions that can be asked here are rather different from those of the operational evaluations that form the focus of this document, but here are some illustrative questions.

4.8.1 Did the preventive action aim to build capacity and infrastructure independently of implementing specific operational crime prevention activities targeted on specific problems? If so:

4.8.1.1 Did the action concern building operational capacity? Innovative capacity?
Improving the operational context for prevention projects?

4.8.1.2 Was the action the establishment of a replicative delivery package, and/or the application of practical process models?

4.8.1.3 Were there indications that the appropriateness, quality, quantity and innovativeness of operational preventive action had increased as an attributable result of the supporting action?

This kind of evaluation requires a clear logic model. The action has to deliver the requisite capacity or infrastructure, and this in turn has to deliver actions (and perhaps organisations, networks and contextual conditions) which are appropriate for local problem and context, and plausible in terms of evidence, theory and practicalities. The effectiveness of the action can be evaluated by a combination of input measures (e.g. were the resources directed to the right problem?),

intermediate output measures (e.g. were the young people given the intended life skills?) and operational outcome measures (e.g. did the boost in life skills then go on to reduce the numbers convicted of crime... and did this lead to a reduction in crime in the area?)