European Crime Prevention Network

Thematic Paper

No. 5

Evaluation of crime prevention initiatives: the principles of evaluation

In the framework of the project 'Towards a European Centre of Expertise on Crime Prevention' - EUCPN Secretariat, June 2013, Brussels

Evaluation of crime prevention initiatives: the principles of evaluation

Abstract

The fifth thematic paper in the series published by the EUCPN Secretariat in collaboration with the Irish Presidency, focuses on the theme of evaluation. It is written in the framework of the third EUCPN Toolbox on the same theme. The toolbox is developed to assist people engaged in evaluation who have limited resources, who often lack the internal expertise to conduct a robust evaluation or who have limited access to information and external support. The aim of this instrument is to develop more insight in 'how and why' to evaluate and to provide a minimum standard of knowledge and skills who are (about to be) involved in programme evaluation of small scale community-based crime prevention initiatives. Attention is paid to the planning of an evaluation, the data collection and analyses, and the communication of results.

This thematic paper aims to be a general introduction to the principles of evaluation based on existing academic literature and provides more detailed information about the second part of the toolbox, the practical guidelines. It is recommended to read the thematic paper in conjunction with the guidelines in the toolbox.

Citation

Legal notice

Authors

Belinda Wijckmans, EUCPN Secretariat, Brussels, Belgium

Barra Ó Dúill, Department of Justice and Equality, Dublin, Ireland

Rosita Vanhauwaert, EUCPN Secretariat, Brussels, Belgium

Contents

Evaluation of crime prevention initiatives: the principles of evaluation

Introduction

This paper gives a general introduction to the principles of evaluation based on existing academic literature. Although it is aiming at providing a broader background, it is closely linked to the second part of the EUCPN Toolbox on evaluation[1], the practical guidelines, following the same structure, going more in-depth on certain issues, etc. Therefore, more than in the previous EUCPN toolboxes, the theoretical and practical part are complementing each other and it is recommended to read this thematic paper in conjunction with the guidelines.

The theme of this paper is explored in three sections on (i) planning an evaluation, (ii) data collection and analysis, and (iii) communication evaluation findings.

Section 1: Planning an evaluation

This section provides a more detailed outline of some of the topics which feature in section 1 of the Practical Guidelines (see p.29) along with a list of reading materials that may prove useful should further exploration of these topics be required.

Topics covered include:

- Evaluation designs

- Literature searching

- Developing evaluation questions

Evaluation types and designs

There are a number of different types of evaluation, which range from evaluations that focus on assessing as to whether a full scale evaluation is feasible to evaluations which examine the longer term impact of programmes. For the purposes of this toolbox we will focus on the two most common types of evaluation those being outcome evaluation and process evaluation.

Outcome evaluation

Outcome evaluations are often the most common form of evaluation called for by a programme's funders. This is primarily because outcome evaluations can help uncover to what extent a programme's objectives have been achieved, i.e. whether a programme actually works. To answer this question, outcome evaluations measure the change that has occurred in a relevant area, for example, anti-social behaviour, rates of reoffending, fear of crime etc, as a result of a programme.

An outcome evaluation examining a programme's attempts to reduce the number of burglaries in an area might focus on:

---

[1] To read the toolbox on evaluation, see the library page of the EUCPN website:
http://www.eucpn.org/library/results.asp?category=32&pubdate

- Changes to people's risk taking behaviour following participating in the programme

- Changes in people's use of security devices

- Changes in people's participation in neighbourhood watch

- Changes in reports of suspicious activity to the police

- Changes in the occurrence of burglaries in a particular area

- Changes in the people's feelings of (in)security

- Whether these changes were more apparent in some sectors of the community than others

- Whether some aspects of the programme were seen as more beneficial than others

In deciding whether a programme has performed well or poorly, some criteria will need to be used, e.g., what level of decrease in reported burglaries would need to be evident in order for the programme to have achieved its target?

Such criteria may be contained in the original programme documentation. If it is not, then criteria should be developed on the basis of what is contained in the relevant literature and in consultation with stakeholders. Furthermore, some knowledge of the situation before the intervention of the programme is also vital for measuring any changes that take place.

Given that outcome evaluations are focused on trying to establish causes rather than describing what is happening, they are mainly quantitative in their approach. In this regard, usually one of three general methods is adapted, all of which try to guard against the extent to which other factors besides the programme are responsible for the change in the area of interest.

Approaches to outcome evaluations

Randomised control trials (Experimental design)

A randomised control trial involves setting up two or more groups and randomly assigning participants to each of these groups. Participants receive the same pre and post participation assessment. With a sufficient number of participants this randomisation process controls for pre-participation differences or other events that may influence the outcome of the trial. It is this ability to control for other possible explanations that has helped establish randomised control trials as a gold standard approach in research design. Nevertheless, these types of trials are complex and require the input of experienced evaluators with the relevant expertise; as such they are costly to design

Intervention and control group (Quasi-experimental design)

The second approach, involves using a control group and an intervention group. In the context of the burglary example (for a full description of the example see Practical Guidelines, p.34), this may mean finding two areas which are similar in size, crime rate

and social composition but only one of which receives the burglary reduction programme. The difference between this design and the previous example is that in this case the participants are selectively chosen. While any changes in relevant areas that occur in the intervention group but not in the control group can be attributed to the programme, the degree to which other possible explanations can be eliminated depends on the context in question. As with the experimental design, this approach can also be complex and costly given that external input from experts will likely be required.

Pre-post design

The last of the approaches involves measuring the occurrence of a phenomenon before the intervention takes place and then measuring again after the programme has been delivered. *All things being equal*, any improvements or changes can then be attributed to the effect of the programme. The extent to which all things are indeed equal is difficult to answer and therefore other possible explanations for change are not easily eliminated. While not as scientifically rigorous as either of the two preceding examples, pre-post designs are simpler and easy to implement and can produce useful results.

Process evaluation

While outcome evaluations seek to establish *whether* a programme works, process evaluations are focused on examining *how* a programme works. Process evaluations assess how a programme is planned, implemented and delivered. Very often this type of evaluation is conducted in order to help verify that a programme is being carried out as originally intended and/or for the purposes of improving programme delivery.

Process evaluations can also be combined with outcome evaluation designs in helping to explain why a programme did or did not achieve its intended impact. This is important from a managerial perspective as it can also help answer the question of whether a programme was a good idea which was poorly implemented or simply just a bad idea. Good ideas which are poorly or incorrectly implemented are unlikely to achieve the intended effects. What's more, such an occurrence may lead to not alone the programme itself being rejected, but also the underlying philosophy of the programme may come to be seen as flawed. It is therefore of great importance that a good understanding of how a programme works in practice be obtained in addition to focusing on the effects of a programme.

There are a number of potential topics that a process evaluation can focus on including:

- To what extent is the programme's intended target group being reached?

- How engaged are service users with the programme (duration and intensity)?

- How satisfied are service users with the programme?

- Are services being delivered as intended?

- Are there sufficient staff numbers to deliver the service?

- Are members of staff appropriately trained to deliver the service?

- What are the members of staff views on the programme?

- To what extent are the relevant stakeholders involved with the programme?

- Have there been any changes to how the programme is implemented?

- Has the programme encountered any obstacles?

In order to answer such questions a wide range of information may be needed, much of which should be collected on an ongoing basis through monitoring. **Monitoring** should, however, be distinguished from evaluation as the former involves the systematic collection of data whereas the latter involves the analysis and interpretation of data.

The kind of information to be collected will depend on the programme in question. However, regardless of area of interest, a literature review involving an examination of the most relevant literature, academic journals and several good practices will provide a good indication of the kind of data to be collected. Stakeholders should also be consulted in this regard.

Examples of the kind of information that could be required may include:

- A description of the service/intervention that the programme provides

- Characteristics of the staff providing the service

- How often the service is provided

- How long the service is provided for

- What participation in the programme involves (intensity)

- Numbers using the service

- Numbers completing the programme

- How stable is service provision

- How satisfied providers and users are with the service

- How closely the original design/plan for the service was followed

- How the programme meets the needs of service users

In deciding whether a programme has performed well or poorly, some criteria will need to be used. This may be contained in the original programme documentation. If it is not, then criteria should be developed on the basis of what is contained in the relevant literature and in consultation with stakeholders.

On the basis of the kind of information needed to answer questions that arise in the context of process evaluations both qualitative and quantitative methods may be used during the course an evaluation. Quantitative methods may be used to measure how many people participated in a programme, the degree to which persons participated, the demographic characteristics of participants and how happy they were with the service. On the other hand, qualitative methods may be used to explore participants and staff views on the programme.[2].

---

2 For more information on evaluation design see: Rossi, P.H., M.W. Lipsey & H.E. Freeman (2007). Evaluation: A systematic approach. Thousand Oaks, CA: Sage.

Searching for relevant literature

Why review the literature?

Conducting a successful evaluation of whatever kind requires that evaluators not only have a good grasp of evaluation methodologies but that they also have an understanding of the relevant topics in the area being evaluated. This requires evaluators to carry out a review of the relevant literature. There are a number of reasons for why this is important.

Firstly, from a very practical perspective it is useful to be able to build on research that has already been done (previous research or evaluations) rather than having to start entirely from scratch. By building on what others have done and thereby seeing what has worked/not worked in the past, costly mistakes can be better avoided.

Secondly, evaluation questions can be made more meaningful as previous research/evaluations may highlight issues that merit further examination.

Thirdly, the findings of other evaluations can help contextualise the results of your own work by providing a backdrop to contextualise findings (Brophy, Snooks & Griffiths, 2008).

Searching for material

At present, thanks mainly to the internet, there is a large body of information available on almost any topic. While this certainly has benefits in terms of making information more easily available, it can also mean that it is much easier to be swamped by information, some of which may not be entirely reliable. It is therefore a good idea to adopt a number of simple methods when conducting a review of the relevant literature.

Firstly, a list of key phrases or terms related to your topic of interest should be developed. These are known as search terms and will determine the information that is found when searching the internet, databases, library catalogues etc. Keep a record of the searches conducted using these terms as well as the results that were found. This will save time in avoiding repeat searches of the same material. *Google Scholar* is a useful online resource in finding academic articles and other publications.

Furthermore, *The Rough Guide to the Internet* (Buckley & Clark, 2009) is also a useful publication in terms of advice on improving internet search techniques.

Secondly, a search can be refined by looking for material published during a particular timeframe, in a specific language or research published concerning a certain country.

Thirdly, for the purposes of ensuring quality control a search can be focused on peer reviewed articles and on particular journals.

Finally, it is always useful to consult, even on an informal, level with persons who have experience in the area being evaluated as this help highlight important information[3].

---

[3] For more information on conducting a literature search see: Brophy, S., H. Snooks & L. Griffiths (2008). Small-scale evaluation in health: A practical guide. London: Sage.

Developing evaluation questions

Why evaluation questions are important

Developing clear and useful evaluation questions is a vital aspect of planning an evaluation. Ultimately, it is the evaluation questions that determine the type of research approach that will be taken during the course of the evaluation (Eck, 2002). The merit of an evaluation question can be judged by how meaningful it is in relation to the programme being assessed and how relevant it is to the concerns of stakeholders (Robson, 2000)**.**

Understanding the programme

Evaluation questions can be developed through reviewing both the programme's background documentation and the relevant literature. This will help highlight any pertinent areas that should be explored. In some cases evaluation questions may exist in a rudimentary form in terms of a project's goals and objectives; however these are often rather vague statements and likely to require further refinement before they may be used as evaluation questions (Maxfield & Babbie, 2010).

It is particularly important, especially in the case of outcome evaluation, that evaluators understand how participation in a programme is supposed to have a particular effect on clients. If there is no clear reason why there should be link between participation in a programme and a particular effect or outcome there would seem to be little point in investigating this in an evaluation.

> *For example, as part of its awareness raising activities, an organisation makes a small number of presentations regarding cyber crime over the course of a year. The organisation then decides that it wants to assess the value of its awareness activities by measuring awareness of cyber crime amongst the general public. Given the level of awareness raising undertaken by the organisation, is it realistic to expect that their efforts increased knowledge of cyber crime amongst the general population? The answer is most likely 'no'. A better way of approaching this issue may have been to examine the extent to which awareness was raised amongst attendees at the presentations.*

In summary:

1. Examine the question in the context of the actual programme activities related to it.

2. Programme components, activities and personnel assignments that relate to programme performance should be identified. Evaluation questions should be formulated in a way that is reasonable given these characteristics.

3. Look at potential questions in the context of findings in relevant areas in the literature.

Developing evaluation questions with stakeholders

In regard to how the views of stakeholders can help in the formulation of effective evaluation questions, it is important to remember that evaluations must generate useable knowledge rather than knowledge for the sake of knowledge (Patton, 1997). It is

not uncommon that evaluation findings go unused. There are a number of reasons for this, such as people not being clear about what information is needed or findings not being as useful as the stakeholders thought they would be at the outset (Rossi, Lipsey & Freeman, 2007).

The likelihood of having unusable findings can be reduced if stakeholders' views are taken into account during the development of evaluation questions. Nevertheless, it must be borne in mind that stakeholders may lack research expertise and may need assistance in developing useful evaluation questions. This process can also help increase positive participation by stakeholders as they will feel they have actively been involved in the design of the evaluation.

The discussions focused on developing evaluation questions may also highlight differing and inconsistent views concerning what a particular project was trying to achieve or how it should operate. In some cases discussions may also indicate that there is a lack of clear understanding regarding what the programme is supposed to achieve. If such issues arise they will need to be addressed prior to the development of the evaluation questions.

How to make sure questions are answerable?

Anyone reading an evaluation report should be able to understand the evaluation questions in terms of what it is they are asking and what it is they are trying to measure.

In order to help ensure that evaluation questions have been formulated appropriately and are answerable the following factors should be considered (Weiss, 1997):

- Identify the group that is to be assessed

- Identify the specific measureable characteristics

- Give example of evaluation findings that might result

- Specify the evaluation criteria (i.e. the threshold that should be reached for success to be achieved, for example, a certain percentage reduction in reoffending rates)

- Have the evaluation sponsors/stakeholders agree that a finding meeting these criteria would answer the question

A useful way to approach the development of such questions is through a process of backward mapping. In other words, decide on the kind of answers that are needed and consider how these may be obtained (Elmore, 1979).

Organising evaluation questions

Once a draft of the evaluation questions have been developed they will need to be arranged by themes. It is likely that a large number of evaluation questions may initially be drafted, many of which may not be used in the end. In selecting the most useful questions focus on the main purpose of the evaluation and what the findings will ultimately be used for[4].

---

[4] For further information on developing evaluation questions see:

Section 2: Data collection and analysis

The following section provides a more detailed outline of topics concerning both quantitative and qualitative approaches to data collection which feature in section 2 of the Practical Guidelines along with a list of reading materials that may prove useful should further exploration of these topics be required.

Topics covered include:

- Issues to consider when using quantitative methods

- Issues to consider when using qualitative methods

Issues to consider when using quantitative methods

Quantitative analysis can vary in complexity from calculating averages to more complicated statistical techniques exploring the relationship between different variables. While it is beyond the scope of this toolbox to explore even a small number of the issues pertaining to quantitative analysis, those covered below provide a very brief introduction into some of the main topics of relevance.

Probability - statistical significance

When evaluators or researchers want to compare two or more groups (people, areas, times) according to some variable of interest they are normally interested in discovering whether the difference between the groups, in terms of this variable, is a real or true difference (i.e. statistically significant) or is simply a product of random chance (i.e. not statistically significant). It should be noted then that 'significance' in the statistical sense does not mean 'important' as its use in everyday speech would seem to imply.

The starting assumption before any analysis is conducted is that there is no real difference between the groups. This is called the *null hypothesis*. A statistical test is then carried out to test the null hypothesis. If the test generates a *p-value* (i.e. probability value) equal to or less than 0.05 the null hypothesis is rejected and the difference between the groups is accepted as being almost certainly true. In statistical terms this means that there is a 95% (or more) probability that the difference found is real compared to a 5% (or less) probability that the difference is simply due to chance.

In calculating statistical significance there is a risk that the null hypothesis is rejected when it is in fact true. This is known as *Type 1 Error*. Conversely, there is also the risk

Davidson, E.J. (2005). Evaluation methodology basics: The nuts and bolts of sound evaluation. Thousand Oaks, CA: Sage.
Maxfield, M. & Babbie, E.R. (2010). Research methods for criminal justice and criminology. Belmont, CA: Wadsworth.
Patton, M.Q. (2001). Qualitative research & evaluation methods. Thousand Oaks, CA: Sage.
Scriven, M. (1991). Evaluation thesaurus (4th ed.). Newbury Park, CA: Sage.
Weiss, C. (1997). Evaluation: Methods for studying programs and policies. Upper Saddle River, N.J.: Prentice Hall.

that the null hypothesis is accepted when it is in fact false. This is known as *Type 2 Error*[5].

Sampling

Sampling is the process of selecting a group of persons from a larger population, carrying out some form of analysis on this group and generalising the results back to the population from which the group was drawn.

Sampling is of concern to those who wish to generalise beyond the group of people directly participating in a study or programme. In most cases small scale evaluations are not interested in generalising beyond a particular programme and therefore issues concerning sampling are usually not a concern. However, in cases where evaluators are seeking to generalise, it is important that the relevant literature and experts working in this field be consulted as sampling is a complex process[6].

Response rate

The response rate of a survey is a measure of how many people actually completed (or in some cases partially completed) the survey when they were approached. The higher the response rate the better. A higher response rate makes it more likely that the results are reflective of the views of the population as a whole rather than a segment of the population who may hold a particular point of view not shared by the majority of the population. If only a small number of those approached respond to a survey, then the results are much more likely to be biased and it would be inappropriate to attribute the findings to the wider population.

Generally, the more interaction the person collecting the data has with the person being surveyed the higher the response rate will be. Therefore, surveys conducted in person or over the phone tend to have higher response rates than postal or internet surveys.

Reliability and validity[7]

Reliability and validity are issues pertaining to the evaluation of measures of concepts. For example, if as part of an evaluation on a programme aimed at improving neighbourhood security, evaluators wanted to measure residents' fear of crime before and after the intervention they would need to think about the reliability and validity of their measure of fear of crime.

Reliability

Reliability refers to the consistency of a measure. This can include:

Stability - If the measure of a concept (e.g., fear of crime) was taken from the same sample on two different occasions and under the same circumstances the results should be highly correlated with each other if the measure is reliable or stable. Conversely, if the results are very different then the measure would seem to be unreliable. However, this may be a result of changes in the sample rather than the measure. This can be particularly problematic if there is a lengthy time lag between the initial and second test.

---

5 For further information see: Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Mahwah, N.J.: Lauwrence Erlbaum Associates.
6 For further information see: Lorh, S.L (2010). Sampling: Design and Analysis. Boston, M.A.: Brooks/Cole.
7 For further information see: Bryman, A. (2012). Social Research Methods. Oxford: Oxford University Press.

Internal reliability – If all of the items (questions) of which the measure (e.g., fear of crime) is made up of are in fact measuring the same thing the measure is said to be consistent. Conversely, if one or more of the items (questions) is measuring something else (e.g., fear of the dark rather than fear of crime) then the measure would be internally unreliable. The most common way to test for internal validity is to use a test known as *Cronbach's Alpha*. This test is available in any common statistical software packages such as SPSS, SAS or R. Scores for Cronbach's Alpha vary between 0 and 1 with a score of 0.8 seen as an acceptable level of internal reliability.

Inter-observer consistency – If there is considerable subjectivity involved in categorising information for example from open ended questions or if there is more than one person involved in the categorisation process then there is a risk that data may be categorised inconsistently.

Validity (construct/measurement)

Validity refers to whether or not the measure we are using is measuring what it is supposed to be measuring and not something else (e.g., is the measure we are using to measure fear of crime really measuring fear of crime or something else?). There are a number of ways of checking validity, including:

Face validity – This involves looking at the questions and using common sense to judge whether the questions are measuring what they are supposed to measure. Other persons with expertise in the area in question should also be consulted when checking face validity.

Criterion validity – This involves taking the results of the measure and comparing them to some previously accepted criterion of what the new measure is suppose to be measuring. For example, if scores from a measure of intelligence were compared with results of school or university exams, it would be expected that those scoring high on the intelligence measure would also have also performed well academically. If this was found not to be the case then it is questionable if the new measure is in fact measuring intelligence and not something else.

Convergent validity – This involves testing the validity of your new measure by comparing it to the results attained from a different measure of the same construct.

Issues to consider when using qualitative research methods[8]

Qualitative approaches to research are far less structured than their quantitative equivalents. While this has the advantage of allowing for more flexibility, this does not mean that there are no quality concerns that need to be considered when using qualitative methods. Many of these issues relate to the reliability of findings stemming from in-depth interviews or participant observation etc. At this basic level issues relating to reliability are somewhat similar to those in quantitative research (see above).

Credibility

Qualitative research often stresses that there are multiple possible accounts of social reality. However, this does not imply that findings from interviews or observations are

8 For further information see: Bryman, A. (2012). Social research methods. Oxford: Oxford University Press.

completely arbitrary. The credibility of the account provided becomes an important criterion for judging the value of qualitative research findings.

There are a number of means by which credibility can be established. Firstly, research should be conducted according to the relevant good practices. Another method is to ask the research participants if the findings reflect what they said or did. Finally, *triangulation* could be done which involves using a different method to access the same data. For example, a researcher might try and confirm what they noticed in an observation by follow-up interviews.

Transferability

The degree to which qualitative findings can be generalised from one location, situation or time to the next is questionable. This is because qualitative research is focused on small samples and detail. Therefore, while qualitative research findings cannot be transferable in the same way as quantitative findings, those conducting in-depth interviews or observations should concentrate on providing as much detail as possible. This has the benefit of providing more information for other researchers working on similar topics.

Dependability

In order to ensure that qualitative findings can be relied upon it is important to keep good records of the research process from start to finish. This might include: field notes, audio recordings, interview guide, list of research participants, research questions, data analysis, etc. This allows others the opportunity to check that proper procedures were followed during the course of the research.

Conformability

Related to the idea of dependability is conformability. This involves the researcher/evaluator being able to confirm his/her findings on the basis of the information collection while at the same time showing that he/she has not allowed personal views to overtly influence research findings or that those findings have not been manipulated so as to conform to theoretical considerations.

Section 3: Communicating evaluation[9]

Effectively communicating the results of an evaluation is of vital importance given that the evaluation's real value lies in being able to act as an effective decision making aid.

Nevertheless, communicating the findings and recommendations of an evaluation can be complicated. The influence that an evaluation can have ranges from directly impacting on policy whereby specific changes are made as a result of an evaluation's findings, to more subtle forms of influence, such as drawing attention to emerging social trends which in turn influences the general direction of policy. The findings of an evaluation can also be used by lobby groups for the purposes of advocating a particular policy direction.

In order to ensure that an evaluation has maximum effect, consideration should be given to the factors that influence how evaluations are received by those who they are aimed at

Presenting the report in a suitable format

The culture and structures of the sponsoring organisation can be an important influence on how an evaluation report is received and acted upon. In this regard consideration should be given to the type of data/information the organisation is most likely to respond to and what kind of information it is likely to ignore (Preskill & Torres, 1999).

Some organisations may take findings presented in a statistical format more seriously as they might see quantitative data as more scientific. However, even if sponsors are interested in quantitative results they may not have an in-depth understanding of statistical techniques and may simply be interested in what the data is saying in layman's terms. In such instances the report should tell the story of the data rather than relying on readers to figure this out for themselves. On the other hand, some organisations may prefer findings of a qualitative nature that sets out views of stakeholders and service users in thematic textual format. Even so, sponsors preferring this approach may not be particularly interested in or aware of qualitative approaches such as grounded theory or phenomenology. Again a report laden with technical jargon is less likely to be effective. This is not to suggest that the findings of a report should necessarily be presented in a simplistic manner as some sponsoring organisations may require more technical reports. Rather the point is that findings should be presented in a format which is appropriate and accessible to the target audience.

Understanding the usability of results

Another important factor to consider is the usability of an evaluation's findings and recommendations. Decision makers are likely to consider the practicalities and feasibility of implementing an evaluation's recommendations in the context of their own organisation in terms of how adapting a particular recommendation is likely to compliment or challenge existing policies. They may also think about what incentives, or indeed disincentives, are there for program staff to enact potential changes stemming from a report's recommendations.

---

9 For further information see: Preskill, H. & R.T. Torres (1999). Evaluative inquiry for learning in organisations. Thousand Oaks, CA: Sage.

It is likely that at least some stakeholders will be reluctant to change their current practices. There may be valid reasons for this or it may simply be a wish to continue with what is familiar. It is therefore important that evaluators provide some guidance as to how the recommendations presented in the report might be feasibly implemented in light of these potential difficulties. Recommendations made in the absence of such considerations may not be implementable in practice. This underlies the importance of including stakeholders in the evaluation process and having a good understanding of the sponsor or target organisation.

Reliability of the findings

The reliability of a report's findings is an important factor in how it is received. In this regard decision makers are likely to pay attention to (a) the quality of the information in the report, including the methodology used, the quality of the data collected, methods of analysis etc, and (b) the extent to which the findings conform with or differ from similar such studies in the area.

This underlies the importance of carrying out well planned, high quality evaluations that can withstand thorough examination. This is particularly the case if the evaluation report contains recommendations that are likely to require significant organisational changes. In such circumstances it is probable that at least some stakeholders may query the validity of the evaluation in terms of the methods used and subsequent the findings. While no evaluation is perfect, those which have rather obvious flaws such as drawing bold conclusions on the basis of questionable or limited data are likely – and rightly so – to be less effective in influencing policy or processes.

References

Alkin, M.C. (2011). *Evaluation essentials: From A to Z.* NY: The Guilford.

Babbie, E.R. (2010). *The practice of social research.* Belmont, CA: Wadsworth.

Bennett, J. (2003). *Evaluation methods in research.* NY: Continuum.

Brophy, S., H. Snooks & L. Griffiths (2008). *Small-scale evaluation in health: A practical guide*. London: Sage.

Bryman, A. (2012). *Social Research Methods.* Oxford: Oxford University Press.

Buckley, P. & D. Clark (2009). *The rough guide to the internet.* London: Rough Guides Ltd.

Clarke, A. (1999). *Evaluation research: An introduction to principles, methods and practice.* London: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, N.J.: Lauwrence Erlbaum Associates.

Connell, P.J, A.C. Kubisch, L.B. Schorr & C.H. Weiss (1995). *New approaches to evaluating communities initiatives: Concepts, methods, and contexts.* Roundtable on comprehensive community initiatives for children and families. Washington, DC: The Aspen Institute.

Dahlberg, L. & C. McCaig (eds.) (2010). *Practical research and evaluation: A start-to-finish guide for practitioners.* London: Sage.

Davidson, E.J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Eck, J.E. (2002). *Evaluating crime prevention. Crime prevention studies. Learning from experience in problem-oriented policing and situational prevention: The positive functions of weak evaluations and the negative functions of strong ones.* Monsey, NY: Criminal Justice Press.

Elmore, R. (1979). Backward mapping: Implementation research and policy decisions. *Political Science Quarterly*, 94(4): 601 - 616.

Fetterman, D.M., S.J. Kaftarian & A. Wandersman (eds.) (1995). *Empowerment evaluation: Knowledge and tools for self-assessment & accountability.* Thousand Oaks, CA: Sage.

Fitzpatrick, J., C. Christie, M.M. Mark (2009). *Evaluation in action: Interviews with expert evaluators.* Thousand Oaks: Sage.

Lorh, S.L (2010). *Sampling: Design and Analysis.* Boston, M.A.: Brooks/Cole.

Love, A.J. (1991). Internal evaluation: Building organizations from within. *Applied Social Research Methods Series*, Vol.24, Newbury Park, CA: Sage.

Mathison, S. (ed.) (2005). *Encyclopedia of evaluation.* Thousand Oaks, CA: Sage.

Maxfield, M. & Babbie, E.R. (2010). *Research methods for criminal justice and criminology*. Belmont, CA: Wadsworth.

Mercier, C. (1997). Participation in stakeholder-based evaluation: A case study. *Evaluation and Programme Planning*, 20(4): 467-475.

Mertens, D.M. & A.T. Wilson (2012). *Program evaluation theory and practice: A comprehensive guide.* NY: The Guilford.

Patton, M.Q. (1997). *Utilization-focused evaluation*. The New Century Text, 3rd. ed. Thousand Oaks, CA: Sage.

Patton, M.Q. (2001). *Qualitative research & evaluation methods.* Thousand Oaks, CA: Sage.

Preskill, H. & R.T. Torres (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Preskill, H. & Russ-Eft, D. (2005). *Building evaluation capacity: 72 activities for teaching and training.* London: Sage.

Robson, C. (2000). *Small-scale evaluation: Principles and practice.* London: Sage.

Rossi, P.H., M.W. Lipsey & H.E. Freeman (2007). *Evaluation: A systematic approach.* Thousand Oaks, CA: Sage.

Royse, D.D., B.A. Thyer & D.K. Padgett (2006). *Program Evaluation: An Introduction.* Belmont, CA: Wadsworth.

Russ-Eft, D. & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change.* NY: Basic Books.

Scriven, M. (1991). *Evaluation thesaurus (4th ed.)*. Newbury Park, CA: Sage.

Scriven, M. (1993). Hard-won lessons in programme evaluation. *New Directions for Programme Evaluation*, No.58. San Francisco, CA: Jossey-Bass.

Shaw, I., J.C. Greene & M.M. Mark (eds.) (2006). *The SAGE handbook of evaluation. Policies, programs and practices.* London: Sage.

Smith, M.F. (1989). *Evaluability assessment: A practical approach.* Boston: Kluwer.

Stern, E. (ed.) (2005). *SAGE benchmarks in social research methods: Evaluation research methods. (Vols. 1-4)*. London: Sage.

Strecher, B & W.A. Davis (1987). *How to focus an evaluation.* Newbury Park, CA: Sage.

Stufflebeam, D.L. & A.J. Shinkfield (2007). *Evaluation theory, models, and applications.* San Francisco, CA: Jossey-Bass.

Tilley, N. & A. Clarke (2006). Evaluation in criminal justice. In: Shaw, I., J. Greene & M. Mark (eds.) *The SAGE handbook of evaluation*. London: Sage.

Weiss, C. (1997). *Evaluation: Methods for studying programs and policies*. Upper Saddle River, N.J.: Prentice Hall.