2011

Improving practice through program evaluation

Renata Phelps

*Southern Cross University*

Publication details

Phelps, R 2011, 'Improving practice through program evaluation', *Centre for Children and Young People: Background Briefing Series*, no. 8.

© Centre for Children and Young People, Southern Cross University, Lismore, NSW, Australia.
Improving Practice Through Program Evaluation
Centre for Children and Young People
Background Briefing Series, no.8

Prepared by Dr Renata Phelps for Southern Cross University’s postgraduate programs in Childhood and Youth Studies
Compiled by Dr Renata Phelps

This Background Briefing has been developed as part of the Graduate Certificate, Graduate Diploma and Master of Childhood and Youth Studies developed and delivered by the School of Education and the Centre for Children and Young People at Southern Cross University.

These awards meet a recognised need, expressed by a range of professionals, for contemporary knowledge and skills to assist them to work more effectively with children, young people and their families.

This and other Background Briefing Papers, together with more information about these awards, can be accessed at www.scu.edu.au/childhoodstudies

SUGGESTED CITATION
Improving Practice Through Program Evaluation: Background Briefing

Introduction..................................................................................................................................................1

Section 1: What do we mean by a program evaluation? ..............................................................................2
  Why is program evaluation important? ........................................................................................................3
  Overcoming the ‘oh no’ factor! .......................................................................................................................4
  Problematising the notion of ‘value’ ..............................................................................................................5
  What constitutes a ‘findings’ and what do we mean by ‘utilisation’? ............................................................7
  What do we mean by ‘impact’, ‘outcomes’ and ‘indicators’? ..........................................................................8
  Program Evaluation: Bright and beautiful, great and small ...........................................................................9

Section 2: Types of program evaluation ...................................................................................................13
  Typologies for program evaluation .............................................................................................................13
  Considering some evaluation approaches and their underpinning ideologies ...........................................16
  Selecting approaches to evaluation ...........................................................................................................22

Section 3: Why involve children and young people in evaluation? ............................................................23

Section 4: Methodology and methods in program evaluation ....................................................................25
  Paradigmatic considerations .......................................................................................................................25
  Methods ......................................................................................................................................................26
  The concept of ‘baseline’ data ....................................................................................................................27
  Sampling ....................................................................................................................................................27
  Ensuring quality and credibility of data and findings ..................................................................................30
  The role of the evaluator ...........................................................................................................................31
  Bringing it all together: Planning for program evaluation .........................................................................32

Section 5: Ethics in Program Evaluation ..................................................................................................33
  Weighing up benefits and risks ..................................................................................................................34
  Access to stakeholders ...............................................................................................................................34
  Informed consent .......................................................................................................................................35
  Privacy, confidentiality and anonymity .......................................................................................................35
  Cultural considerations ...............................................................................................................................36
  Payment to participants ..............................................................................................................................36
  Child Protection ..........................................................................................................................................36

Section 6: Future directions .......................................................................................................................37

References .....................................................................................................................................................38
Inviting your Critical Engagement

Photos used throughout this publication are sourced from Shutterstock (http://www.shutterstock.com). They have been selected to highlight the diversity of ways in which children and childhood can be represented.

We encourage you to engage critically with these images as you reflect on the idea that 'childhood' is socially constructed. Ask yourself, 'What message about children or childhood is being conveyed through this image'? ‘How do these images challenge my understandings of children and childhood’?
Introduction

Most professionals who work with children, young people and their families, whether in educational contexts, family and community services, juvenile justice, counselling or community development (to name but a few), will at some point in their career find themselves involved in program evaluation. As a practitioner you will undoubtedly have asked yourself the questions: ‘Does this program (or intervention) work?'; ‘How do I know?'; ‘What evidence would I look for?'; ‘What aspects could I improve?’. These are essentially evaluation questions and are critical in ensuring the programs and services we provide to children, young people and their families do more good than harm.

For this reason, program evaluation is now a critically important part of most social services in both government and non-government organisations.

Not everyone, of course, finds themselves taking on a major responsibility in conducting an evaluation, but evaluation is now increasingly recognised as integral to program planning and implementation. So understanding what program evaluation is about, how it is conducted and what can help ensure it is successful, is important for any professional.

The first part of this Background Briefing attempts to define some of the key concepts and ideas related to program evaluation. We consider what it is, why we would do it and what forms it can take. In Section 2 we consider types of program evaluation, before discussing briefly how evaluators or organisations might choose which approaches are best applied to their particular program and context.

Section 3 focuses specifically on the importance and value of involving children and young people in program evaluation. Section 4 deals with methodological considerations, including approaches to ensuring the quality and credibility of data and findings. Section 5 deals specifically with ethical issues in program evaluation. Finally, Section 6 considers some future directions and challenges in the field of program evaluation.

Whilst the content provided in this Background Briefing is necessarily introductory in focus, the field of program evaluation is vast and we encourage you throughout this document to build on your existing understandings and practices in conducting robust, relevant and respectful evaluation.
Section 1: What do we mean by program evaluation?

Evaluate refers, literally, to the process of looking at, and judging, the quality or value of something (Bond, Boyd, Rapp, Raphael, & Sizemore, 1997). It is the making of a value judgement of worth or merit based on certain information, and that information generally has to be collected before value judgements can be made.

Of course there are all kinds of things that we evaluate every day in our personal and professional lives, including policies, products and people. In this Background Briefing we are particularly focused on the evaluation of programs. So what do we mean by a ‘program’?

Smith (1989, cited in Owen, 2006) defines a program as a set of planned activities directed toward bringing about specified changes for an identified or identifiable audience. In other words, a program has two elements: (a) a documented plan and (b) action consistent with the plan. Of course programs come in all shapes and sizes and occur within and across disciplinary contexts. At the macro-level a program might be prompted by a government initiative or funding priority and could involve a large-scale response and change across a whole country. At the micro-level, programs are implemented within individual work contexts by teams of people or even individuals. All these forms of programs can be subject to evaluation.

Before moving on from our definition of program evaluation, it is important to attempt to differentiate evaluation from research.

Research and evaluation have a lot in common, particularly when it comes to matters of data collection and analysis and aspects such as ethics, all of which we consider in this Background Briefing. Research, however, is more concerned with advancing the frontiers in a discipline or a field of study (Owen, 2006) – in other words, research is intended to develop ‘new knowledge’, test theories, establish truths and (sometimes) generalise across time and space (Patton, 1997).

Evaluation, in contrast, doesn’t share this imperative... which is not to say that it can’t lead to ‘new knowledge’; rather that the focus is on the need to inform decisions about specific programs, clarify options, identify opportunities for improvement and provide information about programs and policies within particular contexts.

Stakeholders are individuals or groups of people whose interests are ‘at stake’ and can include workers, clients and other beneficiaries, together with other agencies whose services might also be impacted by the program. Stakeholders can also include funding bodies, governments, and the general public – depending on the nature of the program.
Evaluations are often commissioned by a funding body or organisation, which means that evaluators are beholden to some or all of the stakeholders or program administrators and managers (Clarke, 1999; Owen, 2006). That said, research is also sometimes commissioned and can experience similar tensions balancing stakeholders’ interests.

From this we can see that evaluation and research share much in common and hence there can be overlap in the literature related to these two areas of inquiry.

**Why is program evaluation important?**

In differentiating evaluation from research (above) we have highlighted that the emphasis of evaluation is not to prove but to improve. The majority of professionals and practitioners would share an interest in improving practice, both at an individual level as well as at a collective, organisational level. Most of us are continually making judgments and decisions which we think will improve what we do:

> Programme staff routinely reflect on the effectiveness of their practice, asking “Are we doing enough to reach the right people?”; “Are we giving children and families what they need?”; “How can we improve the services that we provide?” Evaluation is merely a way of formalising this reflexive process in order to make better sense of people’s experiences and to capture good practice (Harrington, Lloyd, & Ellison, 2005, p.3).

Formalising evaluation within a program or organisation, and planning systematic and rigorous processes for it, can lead to a greater commitment by staff to act on the evaluation findings, and thus strengthen the likelihood of tangible improvements. However, there are other reasons why organisations engage in program evaluation:

> Increasingly, government, foundations, and other donors are asking managers of nonprofit, community-based organizations for more information about the evaluation and performance measurement of the public services they provide for a range of purposes, including contract monitoring and reporting, organizational learning, and informing public policy and program practice (Carman, 2007, p.60)

Program evaluation can meet expectations of accountability by: assessing progress toward program goals; documenting the quality of your programs and describing the effects on participants; and quantifying the amount of change experienced by program participants (Bond, et al., 1997).

To summarise then, we can identify three very good reasons for program evaluation, namely:

- **capacity building** - in order to build individual and organisational skills and abilities;
- **accountability** - to document outcomes and impacts of programs primarily for external stakeholders; and
- **program improvement** - to enhance delivery and effectiveness of the program (Lambur, 2008).
Overcoming the ‘oh no’ factor!

Evaluation can be very rewarding for those involved in designing, implementing and participating in programs. As Comfort (2010, p.17) explains, it can ‘motivate and help people to see how far they have come, and what they have learned, and to decide their next steps’.

Sometimes, however, staff in organisations can be quite nervous about evaluation. They might know very little about it, they might lack trust in the evaluators or the process, they might have had previous negative experiences of evaluations, they might feel threatened by the ramifications of negative findings, or they might just fail to appreciate the purpose and value of the evaluation (Taut & Alkin, 2003).

Part of the challenge in conducting evaluations, then, is to work to overcome such concerns. Writing in relation to the evaluation of a United Kingdom based program, SureStart, Harrington, Lloyd, and Ellison (2005) acknowledged a range of potential concerns, and illustrate how these could be turned around into more constructive values, as summarised in Table 1.

Throughout this Background Briefing we will consider ideas from the theory and practice of program evaluation which address many of these concerns and can assist you, and/or your colleagues, to make the most of program evaluation in order to support your organisational goals and improve outcomes for children, young people and their families.

Table 1: Concerns about evaluation (adapted from Harrington, et al., 2005)

<table>
<thead>
<tr>
<th>Concern expressed</th>
<th>More constructive view</th>
</tr>
</thead>
<tbody>
<tr>
<td>The purpose of evaluation is to be judgmental and criticise the program.</td>
<td>Evaluation is about learning, not criticising. It reflects the program’s commitment to improvement and sustainability.</td>
</tr>
<tr>
<td>Evaluation will expose weakness and problems and therefore reflect badly on the program and the participants.</td>
<td>Conducting evaluation can result in a great deal of extremely valuable information that can be used to re-shape and develop services.</td>
</tr>
<tr>
<td>Evaluation is complicated and expensive. It is a drain on precious resources of time and money, which could be better spent on service delivery.</td>
<td>Some aspects of evaluation work may be resource intensive but potential benefits of robust evaluation include savings in time and money.</td>
</tr>
<tr>
<td>Evaluation is intrusive and disruptive.</td>
<td>High quality, well-designed evaluation need not be intrusive or disruptive.</td>
</tr>
<tr>
<td>Evaluation will not tell us anything we don’t already know.</td>
<td>Dissemination of evaluation findings can be a useful way of raising key stakeholders’ awareness of the benefits of a Sure Start approach to working.</td>
</tr>
</tbody>
</table>
Problematising the notion of ‘value’

Underlying many (although not all) evaluations is what Owen (2006) calls the ‘logic of evaluation’; namely, a four-fold process of:

- establishing or selecting criteria of worth, as the basis for judgements to be made;
- constructing standards regarding what might be considered acceptable, or how well the program should perform;
- measuring performance and comparing this with the standards; and
- synthesising and integrating the evidence into a judgement of merit or worth.

Underpinning all these elements is the idea of judging or assessing the ‘value’ of a program. In fact, the very notion of evaluation as signaled earlier, implies making a ‘value judgement’.

A distinction might be made here between ‘monitoring’ and ‘evaluating’. As explained by Clarke (1999), monitoring concentrates on examining the processes and procedures involved in delivering a program, with information collected on a regular basis and with a view to making changes.

As such, monitoring is an essential part of evaluation and might generate data for it. However, monitoring is essentially value free. The emphasis is on collecting information without questioning the logic or structure of the program design. Again, the emphasis in evaluation is on making judgements of ‘value’.

So what constitutes ‘value’ and whose values are we referring to?

Before considering these complex questions, let’s explore some related concepts: ‘effectiveness’, ‘independence’ and ‘objectivity’.

Many evaluations seek to make judgements about the effectiveness of a program. At one level, effectiveness refers to whether a program has achieved what it aimed to achieve. However, the emphasis in judgements related to effectiveness can vary considerably.

Schweigert (2006) discusses three meanings of effectiveness, each associated with different evaluative aims and designs. Firstly, effectiveness can be seen as associated with increased understanding of the dynamics of the program, focusing on what elements of the program work well or not so well, and tracking patterns of activity and results throughout time or across sites. A second emphasis relates to accountability – whether a program has achieved its performance expectations, referenced by defined standards or exemplars. Thirdly, effectiveness can be associated with the demonstration of causality – whether a program can be demonstrated to lead to
generalisable outcomes that can be replicated in different contexts (something we discuss in more depth later on). Each of these perspectives on what constitutes ‘effectiveness’ itself embodies a whole range of values and assumptions.

There are two other important and related terms which we should also explore: namely, ‘independence’ and ‘objectivity’. The following distinction between these is made by Markiewicz (2008, p.39):

**Independence** generally refers to the evaluator being awarded the freedom to conduct the evaluation without undue control being exerted by the commissioners of the evaluation, the organization or program delivery personnel.

**Objectivity** refers to the evaluator’s capacity to undertake unbiased and objective assessments and form conclusions during the evaluation.

Many theorists argue that all interpretations or value judgments, and the findings and recommendations that result from them, are subjective, arising from our personal positions, beliefs, experiences, values and orientation in life. From such a perspective, no evaluation can be completely objective or independent. It is often said that value lies in the eye of the beholder – acknowledging that there might be many valid interpretations of the same events, depending on a person’s point of view, interests, and beliefs (House, 2001, p.25). The same set of data can be interpreted in different ways by different people.

For these reasons, program evaluation has been described as a ‘political’ activity (Clarke, 1999). Stakeholders present diverse, varied and often competing interests and anyone involved in evaluation needs to recognise the potential ‘power play’ at work. Evaluators, therefore, play a complex role in gathering together the various views of people who are involved in or with a program (the stakeholders) and juggling these into a format which is acceptable and useful to those interested in the results:

Evaluators are no longer seen solely as experts using scientific research techniques, but rather as navigators of a process of enquiry in a political environment characterised by multiple perspectives, audiences and accountabilities (Markiewicz, 2006).

The view that knowledge is context bound (or ‘situational’ i.e. that it is about concrete events in particular settings), places an increased emphasis on the importance of the views and perspectives of all people who are most directly involved in a program. It also raises important questions about generalisability, the ability to apply knowledge gained in one context directly to another context. As House (2001, p.25-6) describes:

To understand a program one must travel into the program setting in the deeper sense to see where people live and how they think. Their beliefs and judgments should be included in the report even if the evaluator does not agree with them, perhaps especially if the evaluator disagrees. In-depth investigation and reporting honors program participants and gives readers the information to draw their own conclusions. Respect for the local setting and the people striving within it is a central value.

Later in this document we consider matters such as validity and reliability and will again discuss the problematic notion of ‘value’ and ‘judgement’. Firstly, however, let’s clarify what constitutes a ‘finding’ and what happens (or should happen) with findings from evaluations.
What constitutes a ‘finding’ and what do we mean by ‘utilisation’?

The notion of ‘findings’ is central to evaluation, but what this term means and looks like differs greatly between different types of evaluation (Owen, 2006). Findings can encompass:

- **Evidence** – the data and other information which has been collected during the evaluation;
- **Conclusions** – the meanings those involved in the evaluation make from a synthesis of data and information;
- **Judgments** – placing value on the conclusions, whereby criteria are applied in order to state that the program is ‘good’ or ‘bad’ or that results are ‘positive’ or ‘below expectations’ etc;
- **Recommendations** – suggested courses of action, advice to policy-makers, program managers or providers about what to do in light of the evidence and conclusions.

Producing findings is, of course, just one part of the evaluation story. What happens with these findings, and whether they influence practice, is another.

Much has been written about evaluation findings being largely unused and leading to little change in the nature and delivery of programs (see for example, Patton, 1997). Evaluators often complain that their findings are ignored, while decision makers lament that the results of an evaluation are disappointing and do not tell them things they need to know. So what might lead to an evaluation being viewed more positively and having an influence on practice? Here we need to discuss the concept of ‘utilisation’.

Owen (2006) identifies two distinct meanings of utilisation. Firstly, it can refer to ‘instrumental use’ which is where an evaluation directly affects decision-making and in some cases the program itself. A second and broader understanding of utilisation includes ‘conceptual use’ or enlightenment. This applies where an evaluation influences thinking about a program, but does not lead directly to change. Over time, evaluations can cumulatively affect decision making in relation, for example, to policies or broader practices.

The utilisation of an evaluation can be influenced by both the characteristics of the evaluation and the characteristics of the program setting. Cousins and Leithwood (1986, cited in Owen, 2006) identify six factors which affect utilisation, namely:

- **Relevance** – degree to which the evaluation focused on issues relevant to the intended audience;
- **Credibility** – whether the evaluator is credible and believable;
• **Quality** – the quality of the data informing the evaluation, including choice of methods and rigor of collection and analysis;

• **Findings** – the nature of the results and the implications for decision making and the relationship to the concerns of the audience; whether findings are practical; and the extent to which findings conflict with current policies and practices;

• **Communication** – quality and quantity of communication during the evaluation;

• **Timeliness of reporting** – whether the evaluation is complete in time to meet the decision needs;

• **Commitment** – attitudes toward the role of evaluation in program planning and policy change;

• **Information needs** – the perceived need for relevant knowledge and the types of information that audience turn to for decision making and problem solving;

• **Personal characteristics** – the attitude of individuals toward evaluation;

• **Decision making** – the context in which decisions, both retrospective and prospective are made;

• **Political climate** – whether findings are consistent with existing political realities in the organization; and

• **Financial climate** – current level of support for the program.

**What do we mean by ‘impact’, ‘outcomes’ and ‘indicators’?**

As we will read later in this paper, the emphasis of an evaluation can be variously balanced between identifying what has been achieved and understanding how it has been achieved. The first is concerned with outcomes and the second with process (Green & South, 2006).

The terms ‘outcome’ and ‘impact’ are sometimes used interchangeably, although some distinctions can be made. ‘Impact’ generally refers to more immediate effects of the program while ‘outcome’ refers to longer-term effects (Green & South, 2006).

Another term which you will often encounter when engaging with the evaluation literature is ‘indicator’. Again, this term can mean different things in different contexts.

Some indicators are ‘big picture’, and quite generic, while other indicators can be quite specific. For example, Education Scotland has developed a set of quality indicators which they advocate being used across diverse organisations and programs to evaluate services for children and young people (Crotty, 1989). These indicators stem from six high-level questions or themes, namely:

- What key outcomes have we achieved?
- How well do we meet the needs of our stakeholders?
- How good is our delivery of services for children and young people?
- How good is our management?
- How good is our leadership?
- What is our capacity for improvement?
If we take, for example, the third of these themes, one of their quality indicators is ‘Impact on children and young people’ from which they identify two indicators, namely:

- The extent to which children and young people are safe, nurtured, healthy, achieving, active, respected and responsible, and included; and
- The extent to which children, young people and their families report that services are enabling children and young people to become safe, nurtured, healthy, achieving, active, respected and responsible, and included.

Another example is provided by Newell and Graham (2010b), who identify 12 recurring policy themes related to child and family services within the current Australian social policy context. These issues pertain to: the nature of the programs being delivered; who is and is not being reached and/or engaged by them; how well the programs are meeting the needs of those accessing them; the service providers’ capacity and confidence to deliver the programs; different client groups’ levels of progress in relation to a range of relevant short-, medium- and long-term outcomes; and the nature and effectiveness of inter-agency functioning. These 12 themes signal a number of issues to be addressed in order to provide both funders and fundees with a comprehensive range of evaluation information.

Indicators can be developed within organisations or they can be set by funding bodies, including governments. Sometimes they relate more broadly to accepted standards in a particular context or country and other times they are established by professional bodies or agencies. They can take the form of long-term outcome indicators or short-term impact indicators; and they can vary as to whether they are intended to provide guidance to an evaluation or whether they are strict criteria which must be demonstrated to have been met.

**Program evaluation: Bright and beautiful, great and small**

Already you will be gaining an understanding that program evaluation is a very complicated and complex arena. There are many understandings of what constitutes ‘good’ evaluation and what it means to ‘judge’ the value of a program. These understandings have changed and evolved over time and, as with any field of professional practice, there are ‘trends’ and ‘fashions’, where particular types of evaluation become popular and others less so. Later in this Background Briefing we consider in more detail some of the many types of evaluation, but before we do so we need to explore some more fundamental ideas about the nature and characteristics of evaluation.
Formative and summative evaluation

The first distinction which might be drawn is the difference between formative and summative evaluation. **Formative evaluation** helps to ascertain how a program is going while it is still being developed or is in progress, helping to make changes, identify issues or correct problems before it’s too late (Bond, et al., 1997). It can also help shape and refine data collection activities, which will contribute to summative evaluation.

**Summative evaluation**, then, occurs when you are summing up what has been achieved at the end of the program, or at appropriate point(s) during the implementation of an ongoing or multi-year program (Bond, et al., 1997).

The distinction between formative and summative evaluation is important, however these activities aren’t always discrete and, as argued by Bond et al., both activities are part of ongoing evaluative processes:

*Data collected while the program is in progress can be used in the summative evaluation to gauge impact. Similarly, information collected at the end of a program can be used in a formative way for designing an improved or expanded program or new programs with similar goals (Bond, et al., 1997, p.6).*

What is evident, then, is that **evaluation shouldn’t be viewed as something that happens only at the end of a program**. Ideally, evaluation is designed, and ‘built in’, right from the start.

**Effective evaluation is not an “event” that occurs at the end of a project, but is an ongoing process which helps decision makers better understand the project; how it is impacting participants, partner agencies and the community; and how it is being influenced/impacted by both internal and external factors.** Thinking of evaluation tools in this way allows you to collect and analyze important data for decision making throughout the life of a project: from assessing community needs prior to designing a project, to making connections between project activities and intended outcomes, to making mid-course changes in program design, to providing evidence to funders that yours is an effort worth supporting (WK Kellogg Foundation, 1998, p.3).

Thus, from the time you begin to consider implementing a program you are clarifying what needs you are trying to address, who your audience will be, and the setting, or context, in which your program will operate (Bond, et al., 1997), and these are all important preliminary steps in the evaluation process.

Of course, while the ideal might be for evaluation to be embedded throughout a program’s lifespan, for various reasons this isn’t always the reality in practice. Sometimes evaluation happens in less than ideal situations, where there are constraints on budget, time, and/or data. As Bamberger et al. (2004) explain, sometimes a project is already well advanced before evaluation is prioritised and there is limited to no baseline data available.
Internal and external evaluations

If evaluation is something that should, ideally, happen all the way through the implementation of a program, then it becomes evident that evaluation is not (always) just something done by external evaluators. Rather, evaluation can be something that all members of an organisation are involved in. Of course, some very large organisations have their own evaluation units, with dedicated staff, and many evaluations are conducted by contracting in ‘outsiders’.

There are advantages and disadvantages of using both internal or external evaluators and some of these are summarised in Table 2.

Ultimately, the structure and size of the program, and the organisation itself, affects their ability to support evaluation of their programs and activities (Lambur, 2008). Externally conducted evaluations aren’t a reality for many small community organisations. We will read much more about forms of evaluation which involve differing levels of participation by staff and other stakeholders later in this Background Briefing.

Scope and scale

A further distinction which might be drawn between different types of evaluations relates to scope and scale. Often very large scale programs are implemented, for example, across a whole state or country, or even internationally. The success of such programs will be influenced by a whole range of complex factors. There is generally a recognised benefit in evaluating such programs overall, but also a need to look much more closely at the local level and what influences the success or other of the program within local contexts. In such instances, ‘dual level’ evaluation often occurs, where evaluation is carried out at both a national and local level (Allen & Black, 2006).

Table 2: Advantages and disadvantages of internal and external evaluators (adapted from Clarke, 1999, p.23)

<table>
<thead>
<tr>
<th></th>
<th>Internal evaluators</th>
<th>External evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>Internal evaluators will be:</td>
<td>External evaluators have:</td>
</tr>
<tr>
<td></td>
<td>• familiar with the history, background, policies, issues and culture of the</td>
<td>• an independent stance and offer a fresh</td>
</tr>
<tr>
<td></td>
<td>organization;</td>
<td>perspective;</td>
</tr>
<tr>
<td></td>
<td>• likely to be more committed to implementing</td>
<td>• an objective, critical approach;</td>
</tr>
<tr>
<td></td>
<td>evaluation recommendations, having been</td>
<td>• an overview of numerous organisations to</td>
</tr>
<tr>
<td></td>
<td>responsible for producing them;</td>
<td>serve as comparisons;</td>
</tr>
<tr>
<td></td>
<td>• likely to focus on the central concerns as</td>
<td>• a knowledge and experience of a wide range</td>
</tr>
<tr>
<td></td>
<td>perceived by management.</td>
<td>of evaluation techniques;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• a resilience to intimidation by management.</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Internal evaluators may:</td>
<td>External evaluators may be:</td>
</tr>
<tr>
<td></td>
<td>• have a vested interest in a particular</td>
<td>• ignorant of internal matters so that</td>
</tr>
<tr>
<td></td>
<td>outcome;</td>
<td>judgments may not reflect the complex</td>
</tr>
<tr>
<td></td>
<td>• often be over-influenced by the history and</td>
<td>reality of the situation;</td>
</tr>
<tr>
<td></td>
<td>knowledge of organizational issues;</td>
<td>• unaware as to who are the key players in a</td>
</tr>
<tr>
<td></td>
<td>• sometimes be over-influenced by the known</td>
<td>particular setting and thus more easily misled by interested parties;</td>
</tr>
<tr>
<td></td>
<td>views of management;</td>
<td>• more interested in a report than its</td>
</tr>
<tr>
<td></td>
<td>• be unlikely to have had experience of a broad</td>
<td>implementation;</td>
</tr>
<tr>
<td></td>
<td>range of evaluation techniques;</td>
<td>• influenced by the need to secure future</td>
</tr>
<tr>
<td></td>
<td>• be less committed to the need for evaluation;</td>
<td>contracts;</td>
</tr>
<tr>
<td></td>
<td>• be inclined to favour programs developed</td>
<td>• insensitive to organisational norms and</td>
</tr>
<tr>
<td></td>
<td>within their own unit or section;</td>
<td>internal relationships;</td>
</tr>
<tr>
<td></td>
<td>• find it difficult to encourage stakeholders in</td>
<td>• primarily responsible to an external</td>
</tr>
<tr>
<td></td>
<td>their own organisation to actively participate in the evaluation process.</td>
<td>organisation.</td>
</tr>
</tbody>
</table>
An Australian example of such an approach might be the evaluation of the ‘Communities for Children’ initiative, a component of the Australian Government’s Stronger Families and Communities Strategy 2004–2009, managed by the Department of Families, Community Services and Indigenous Affairs. This program was designed to implement the key action areas in the National Agenda for Early Childhood, namely promoting healthy young families; early learning and care; supporting families and parenting; and creating child friendly communities (Cortis, 2008). The program funded a whole range of activities that differed between various Communities for Children sites around Australia. These activities were guided by local needs and were intended to build on (rather than replicate) existing infrastructure. Local focus programs included early learning and literacy initiatives, nutrition programs and community-building events.

The initial implementation of ‘Communities for Children’ was evaluated at the national level and various local evaluations also occurred, focused on specific initiatives. At the National level, a National Evaluation Consortium was formed, comprising the Social Policy Research Centre (SPRC) at the University of NSW, the Australian Institute of Family Studies (AIFS) and other research advisers (Edwards, et al., 2009; Social Policy Research Centre, 2007). The Centre for Children and Young People was involved in a number of the evaluations on the North Coast of NSW (see Newell, 2008; Newell & Graham, 2009a, 2009b) and reports on other local evaluations are also accessible (for example, Centre for Community Child Health - Royal Children’s Hospital & Broadmeadows Early Years Partnership, 2007; Turner, 2008).

Relative emphasis on process and outcomes, improvement and assessment

The emphasis in evaluations can also vary between improvement and assessment and between process and outcome. The distinctions between these can be summarised as in Table 3 (derived from Chen, 1996 in Clarke, 1999, p.12).

<table>
<thead>
<tr>
<th>Program stages</th>
<th>Evaluation function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improvement</td>
</tr>
<tr>
<td>Process</td>
<td>Process-improvement evaluation – focus on detecting strengths and weaknesses in a program with a view to altering or adjusting implementation</td>
</tr>
<tr>
<td>Outcome</td>
<td>Outcome-improvement evaluation – concentrates on how the parts of a program affect the program outcomes, taking implementation issues into account, but there is no formal assessment of the program as a whole</td>
</tr>
</tbody>
</table>

Summary

The message we might take from this section, then, is that program evaluation can take many forms ‘bright and beautiful, great and small’! It can be something that happens at the end of a program’s lifespan, or ideally can occur concurrently with the program, influencing its development and refinement. Evaluation can be conducted by external ‘experts’ or can be something driven by program participants themselves. It can occur at national and international levels and also focus on very local initiatives. It can make overall judgements about a program, or focus more on how a program can be improved. Evaluation is something that, in one form or another, is part of the everyday practices of all professionals.
Section 2: Types of program evaluation

In this section we consider the wide range of approaches to program evaluation and the different ways that evaluations can be understood and classified. As we will see, the choices around which types of evaluations are implemented are often influenced by matters of ideology and beliefs about what is ‘important’ in evaluation, as well as the purpose of the evaluation and the intended audience and what they expect.

Typologies for program evaluation

Many writers in the area of program evaluation have proposed frameworks which we can use to consider different types and forms of program evaluation. For example, in the late 1980s, Darren Caulley, a well-known academic from La Trobe University, proposed the following framework of ten dominant program evaluation models (Caulley, 1989):

1. **Measurement model**: emphasis is placed on assigning numbers to objects and events according to rules and these numerical indicators (or performance indicators) are used to demonstrate quality.

2. **Blue-ribbon panel model**: includes such variants as Royal Commissions, Committees of Inquiry and Review Panels and is a model favoured by governments and politicians. Often proceeded by a public or media outcry and normally with the purpose of reform. Emphasis is placed on the crediblility of the panel members.

3. **Management model**: key criteria are effectiveness, efficiency and accountability and there is an emphasis on inputs and outputs.

4. **Objectives model**: the key criteria here is the extent to which a program has achieved its objectives.

5. **Decision-making model**: evaluation is defined as a process of delineating, obtaining and providing useful information for judging decision alternatives and value judgements are made in making those decisions. This model was originated by Daniel Stufflebeam.

6. **Experimental model**: makes use of experimental and quasi-experimental design and an emphasis is on causality i.e. whether it is the program that is directly having an effect.
7. **Needs model:** Focuses on whether the program is satisfying clients needs. While needs are distinguished from wants however such decisions are recognised as value-laden and context-bound. decisions here about.

8. **Expert model:** An acknowledged expert in the area is brought in to make judgements and criticisms about the program (sometimes termed consultants, inspectors, critics or connoisseurs). The expert’s values are predominant.

9. **Democratic model:** Proponents of this model see previous models as authoritarian and/or bureaucratic and as assuming value consensus. Democratic models emphasise power sharing and value pluralism. The evaluator consults various stakeholders (staff, participants, funders etc) and aims to discover their different perspective. There are various models of democratic evaluation, including the responsive evaluation approach of Robert Stake. We discuss these approaches in more detail late in this Background Briefing.

10. **Self-evaluation model:** Carried out by those directly involved in the program itself, whether that be a single person, a group, an institution or an organisation. Reflection on the program’s goals, problems, needs and values is emphasised. The purpose is generally change, reform or improvement and such change is more likely through self-evaluation. Action research (which we will consider later in this Background Briefing) can be used as a form of self-evaluation.

Caulley’s framework is just one of many. For instance, Stufflebeam (2000b) has produced a taxonomy of 22 different approaches to program evaluation and his work is widely cited. Another useful framework is presented by Owen (2006) who identifies five categories or ‘forms’, the purposes and orientations, typical issues, major approaches, assumptions and imperatives of which are summarised in Table 4.

Frequently, however, program evaluations contain elements of multiple types of evaluation approaches. Such **multi-modal or multi-method approaches** can “help in better understanding the needs of stakeholders and program recipients, producing better method designs, and yielding more accurate recommendations by which to enhance program development and change” (Bledsoe & Graham, 2005, p.303). Multimodal evaluation might include multiple perspectives or ideologies, multiple disciplinary perspectives, multiple methods and multiple levels throughout an organisations. While we might consider multimodal approaches as representing a kind of model in their own right (Scriven, 2000), evidently they draw from the philosophical and practical features of each individual approach.
<table>
<thead>
<tr>
<th>Evaluation Form</th>
<th>Orientation</th>
<th>Typical issues</th>
<th>Key approaches</th>
<th>Assumptions</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proactive</td>
<td>Synthesis</td>
<td>• Is there a need for the program?</td>
<td>• Needs assessment</td>
<td>What is already known should influence action</td>
<td>Importance of external frame of reference</td>
</tr>
<tr>
<td>evaluation</td>
<td></td>
<td>• What do we know about this problem that the program will address?</td>
<td>• Research synthesis (evidence-based practice)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What is recognised as best practice in this area?</td>
<td>• Review of best practice (Benchmarking)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Have there been other attempts to find solutions to this problem?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What does the relevant research or conventional wisdom tell us about this</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>program?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What do we know about the problem that the program will address?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What could we find out from external sources to rejuvenate an existing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>policy or program?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarification</td>
<td>Clarification</td>
<td>• What are the intended outcomes and how is the program designed to achieve</td>
<td>• Evaluability assessment</td>
<td>Program rationale and design needs to be laid out</td>
<td>Importance of making intervention explicit</td>
</tr>
<tr>
<td>evaluation</td>
<td></td>
<td>them?</td>
<td>• Logic development</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What is the underlying rationale for this program?</td>
<td>• Ex-ante</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What program elements need to be modified in order to maximise the intended</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>outcomes?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is the program plausible?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Which aspects of this program are amenable to a subsequent monitoring or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>impact assessment?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td>Improvement</td>
<td>• What is this program trying to achieve?</td>
<td>• Responsive</td>
<td>Those close to action need information for ongoing change</td>
<td>Importance of provider involvement</td>
</tr>
<tr>
<td>evaluation</td>
<td></td>
<td>• How is this service going?</td>
<td>• Action research</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is the delivery working?</td>
<td>• Developmental</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is delivery consistent with the program plan?</td>
<td>• Empowerment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How could delivery be changed to make it more effective?</td>
<td>• Quality review</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How could this organisation be changed so as to make it more effective?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring</td>
<td>Checking/</td>
<td>• Is the program reaching the target population?</td>
<td>• Component analysis</td>
<td>Programs need to be monitored to ensure quality</td>
<td>Importance of quality control</td>
</tr>
<tr>
<td>evaluation</td>
<td>refining/</td>
<td>• Is implementation meeting program benchmarks?</td>
<td>• Devolved performance assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>accountability</td>
<td>• How is implementation going between sites?</td>
<td>• Systems analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How is implementation now compared with a month ago?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Are our costs rising or falling?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How can we fine-tune the program to make it more effective?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is there a program site which needs attention to ensure more effective</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>delivery?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact</td>
<td>Learning/</td>
<td>• Has the program been implemented as planned?</td>
<td>• Objectives-based</td>
<td>Need to know what works and why</td>
<td>Importance of transferability: contribution to</td>
</tr>
<tr>
<td>evaluation</td>
<td>accountability</td>
<td>• Have the stated goals of the program been achieved?</td>
<td>• Needs-based</td>
<td></td>
<td>funded knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Have the needs of those served by the program been achieved?</td>
<td>• Goal-free</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What are the unintended outcomes?</td>
<td>• Process-outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Does the implementation strategy lead to intended outcomes?</td>
<td>• Realistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How do differences in implementation affect program outcomes?</td>
<td>• Performance audit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Is the program more effective for some participants than for others?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Has the program been cost-effective?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Considering some evaluation approaches and their underpinning ideologies

There isn’t scope within this document to deal in depth with all types of evaluation. However, in the following sections we have chosen to focus on some specific forms, either because they are particularly ‘topical’ in educational or social and community contexts, or because they represent part of the more contemporary debate in relation to program evaluation. Specifically, we focus on:

- Randomised control trials
- Outcome evaluation
- Participatory approaches to evaluation
- Responsive evaluation
- Utilisation-focused evaluation
- Empowerment evaluation
- Practical Participatory Evaluation

Note that we are not claiming that these categories are mutually exclusive. In fact, the opposite is true. With the exception of randomised-control trials and outcome evaluation, most of the approaches listed might be considered ‘participatory’. However, each brings with it particular nuances which warrant discussion.

Further, while we call these ‘types’ of evaluation, they might also be recognised as representing particular evaluation “ideologies” rather than “methods” or “approaches”, since they really concern the values and principles which practitioners believe should dominate and direct evaluative practice (Smith, 2007). Later in this paper we return to discussion of the notion of ‘ideology’.

Randomised control trials

Traditionally, the ‘gold standard’ of research and evaluation has been considered to be the randomised control trial and this approach has come to dominate in some countries, contexts and disciplines. This is an approach which is experimental in nature in that it occurs under controlled conditions and the aim is to show causal relationships between certain outcomes and the “treatments” or services aimed at producing these outcomes.

It is referred to as ‘randomised’ because each individual or case has an equal likelihood of being placed in a ‘treatment group’ or a ‘control group’ i.e. of being involved in the intervention being evaluated or not being involved. Baseline measures are carried out for both groups before the intervention is applied to the treatment group. The focus is placed on measuring and comparing the outcomes for those in the two groups, based on statistically significant changes in certain outcomes, and every effort is made to either minimise, or account for, unrelated variables that might influence the outcomes (Clarke, 1999; WK Kellogg Foundation, 1998). A diagrammatic representation of the process is provided in Figure 1.

Figure 1: Randomised control trial design (from Green & South, 2006)

As Smith (2007, p.171) emphasises:

Advocates of randomized, controlled trials... follow a fairly clear set of principles about the proper function of evaluation, who should control resources and information, who should participate and how, and what is to be considered acceptable evaluation practice.
More recently, the randomised control-trial approach has been critiqued for a number of reasons. It is seen to be very costly in many circumstances, and some believe that there are areas of inquiry that are simply not amenable to the approach because various considerations prevent the selection of a truly random group of participants (Comfort, 2010). Another limitation, which we can’t fully do justice to here, is that the criteria necessary to conduct such evaluations limit their usefulness in the social sciences and that they are ill equipped to help us understand complex community initiatives. Further, the capacity of such approaches to inform change is limited:

Clearly, demonstrating effectiveness and measuring impact are important and valuable; yet we believe that it is equally important to focus on gathering and analyzing data which will help us improve our social initiatives. In fact, when the balance is shifted too far to a focus on measuring statistically significant changes in quantifiable outcomes, we miss important parts of the picture. This ultimately hinders our ability to understand the richness and complexity of contemporary human-services programs—especially the system change reform and comprehensive community initiatives which many of you are attempting to implement (WK Kellogg Foundation, 1998).

These and other issues have been addressed in some detail by Bickman and Reich (2009).

None-the-less, randomised control trials continue to be viewed (rightly or wrongly) by many individuals and organisations as the optimum approach to evaluation.

**Outcome evaluation**

Outcome evaluation is primarily focused on measuring the effects or results of programs, rather than their inputs or processes and generally involves judgements against an established standard (Kellaghan & Madaus, 2000). Outcome evaluation doesn’t usually seek to describe or specify what is actually happening in the program or to account for the complex factors influencing the outcome...it is, in most cases, just focused on the outcome.

*What matters is the difference you make, the results you achieve. It is insufficient to argue that you work hard, or that you care a lot, or that people enjoy coming, or even that you are extremely important – the point is whether or not positive outcomes result (Comfort, 2010, p.41).*

Judgements can be explicit or implicit (for example, where the achievements of students might be compared to a norm and it is left for the public to make the evaluative judgement).

Kellaghan & Madaus (2000) explain that outcome evaluation has been growing in popularity for a range of reasons, including due to a rise in ‘managerialism’ (with a focus on deliverables and accountability), an increased influence of accounting and auditing within public administration, and a perception that other evaluation approaches are costly, slow and complex without paying sufficient attention to outcomes. We see many of these factors playing out in educational contexts, where governments seek to hold schools accountable for students’ outcomes.
Outcome evaluation can be beneficial where a program has well-articulated goals and it can focus organisational attention on achieving those outcomes. However, there are a number of identified shortcomings to the approach. It is less likely to lead to greater understanding of what goes on in programs, the factors that affect achievement of outcomes and who is responsible where roles are unclear (Kellaghan & Madaus, 2000). Further, because the evaluation is focused on very specific and limited outcomes it may not reflect the broader achievements of a program. Another issue is that organisations and professionals can be tempted to chase targets, in some instances avoiding cases where it is not so easy to show positive outcomes (Comfort, 2010). For these reasons, some argue that outcome evaluation is less valuable in supporting continual improvement.

**Utilisation-focused evaluation**

This ‘school’ of evaluation owes much to the work of Patton (1997, 2002b). It is based on the premise that evaluations should be judged by their utility and actual use, and that such considerations should be foremost in the evaluation design, from beginning to end.

*Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilisation-focused evaluation is on intended use by intended users (Patton, 1997, p.20).*

The evaluator facilitates judgements and decision making by intended users, rather than being a distant and independent judge themselves. There is a critical awareness of whose values frame the evaluation and there are close working relationships and lots of negotiation.

Utilisation-focused evaluation does not advocate particular content, models, methods or theories but rather is a process for selecting the most appropriate of these for the particular situation.

Patton (2011) has most recently expanded these ideas to his discussion of ‘developmental evaluation’, which we allude to in the final Section of this Background Briefing. Developmental evaluation is particularly targeted toward social innovators and others working to bring about major change.

**Responsive evaluation**

Like utilisation-focused evaluation, responsive evaluation (Stake, 2000) emphasises the usefulness of the findings to people in and around the program. An evaluation is viewed as “responsive” if it is oriented to program activities, responds to audience requirements for information, and refers to different value perspectives in reporting the successes and failures of the program.

Rather than being driven by program objectives or hypotheses as advanced organisers, responsive evaluation focuses more on issues. After talking to various stakeholders to become acquainted with a program, the evaluator acknowledges certain issues or potential problems and these provide a structure for continuing discussions and data gathering. This approach means that it is not just issues that are foreshadowed at the beginning of an evaluation that gain focus. Rather, issues that emerge during the evaluation are of equal importance. Responsive evaluation is most likely to involve multiple methods for collecting data.

An important consideration in responsive evaluation is that reporting is consistent with the ways in which people assimilate information and arrive at understanding and, particularly, on providing opportunity for facilitating vicarious experience and conveying a holistic impression of the complexity of the program (Stake, 2000).
Participatory approaches to evaluation

One of the most important distinctions between different types and approaches to evaluation is whether (and to what degree) they are participatory. Participatory approaches to evaluation emphasise the direct and active involvement of various stakeholders.

Such approaches are grounded in ‘constructivist’ philosophy - an understanding that human beings are active interpreters and ‘constructors’ of their world and that meanings are not set or given but rather are shaped through interactions. In this sense, ‘meaning construction is a dialogical process; subject and object mutually influence each other and will change as a result of this process. This means that evaluators cannot be objective’ (Abma, 2006, p.32). We discuss these matters again later in this paper.

Hasenfeld, Hill and Weaver (2002) identify twelve principles of participatory evaluation, which are paraphrased here as:

1. The views and interests of all stakeholders should be articulated, understood, and taken into account;
2. Evaluators should understand that programs continue to change in response to external and internal pressures;
3. Rather than taking a program’s design as given, evaluators should focus on the actual service delivery system as it is enacted by staff and experienced by participants;
4. Evaluators should draw on the whole range of research methods available, using different methodologies when appropriate to the evaluation;
5. Evaluators should not merely be neutral recorders of program processes and outcomes but catalysts for organisational change, providing regular feedback, facilitating discussion of findings, and offering alternative solutions to issues raised;
6. Evaluators should partner with program managers and staff and involve them in the evaluation design and implementation;
7. Evaluation can foster a “learning organization”, in which staff continue to evaluate their performance after the evaluators leave;
8. Development and refinement of management information systems is crucial to the evaluation, and provides key learning opportunities for program managers;
9. The trust established and nurtured between the evaluators and all stakeholders is critical for a successful evaluation;
10. Protecting the rights, and respecting the dignity, of everyone involved, especially through confidentiality, is essential to establishing and maintaining trust;
11. Whether formal or informal, frequent information sharing sessions among participants is central;
12. Evaluators should be experienced in facilitating program changes and have the sensitivity and skill to develop and maintain trusting relationships with a range of stakeholders in an ever-changing program.
The degree to which an evaluation is ‘truly’ participatory can vary along a continuum. At one end stakeholders’ views are sought and taken into account, but at the other end of the continuum stakeholders might also be involved in the evaluation design and conduct – they might formulate questions, select participants, collect data and even interpret findings.

This form of evaluation is very consistent with participatory action research and Lennie (2005) claims that participatory evaluation methodologies emerged from the extension of participatory action research to the field of evaluation. Some evaluators specifically use the cyclical process which characterises action research, with participants collecting and sharing evidence, reflecting on this evidence to determine what works and what doesn’t, refining programs and reformulating policy, and then (again) collecting and sharing evidence (for example, Batterbury, n.d.).

The cyclical approach of an action research evaluation gives groups the chance to progressively look at and think about whether the project is doing what it set out to do and if there is anything that could be done better, and to report on this regularly enough to justify changes if they are needed. If the organisation is already geared to continuous improvement, incorporating the cycles of an action research evaluation and writing up what happens may only be a small change to the way things are done. However if it is a new approach, it may take longer to get used to. Because an action research evaluation is participatory and happens in repeated cycles while the project is happening, it often takes longer than other kinds of evaluations (Lienert, 2002).

As Smith and Champagne (Smits & Champagne, 2008) explain, participatory evaluation can be further categorised according to its ultimate aim. When a major goal is to generate more social justice for disenfranchised minority groups, it is called ‘empowerment evaluation’ and when the participative process aims to increase the use of evaluation results through the involvement of intended users, it is known as practical participatory evaluation (Smits & Champagne, 2008). We will consider these two approaches now.

Practical participatory evaluation

Practical participatory evaluation is characterised by dialogue and the diversity and role of stakeholders in the process. The approach acknowledges cognitive, affective, and political dimensions, including participants’ feelings, and the process contributes to a learning phase that reinforces understandings and the ownership of results and, eventually, a greater sense of obligation to follow through on the results (Smits & Champagne, 2008). Evaluators and stakeholders interact in a partnership and there is stakeholder involvement in evaluative decision making (Smits & Champagne, 2008).

Smit and Champagne (2008) propose a framework of four key concepts which they believe constitute practical participatory evaluation, namely:

- **Interactive data production**, such that data are collected in collaboration between evaluators and practitioners, taking into account the values of each, including scientific rigor, relevance, and feasibility within the given the context.

- **Knowledge co-construction** whereby data are transformed into co-constructed knowledge in a reflexive and dialogic way.

- **Local context of action** acknowledges the environmental conditions and the wide range of interacting variables that affect the action, including the initial conditions that could affect the evaluation process as well as other emergent factors such as the arrival of a new stakeholder.

- **Instrumental use** is the process by which potential knowledge is used to make decisions on action. This knowledge is called actionable knowledge (Smits & Champagne, 2008).
Empowerment evaluation

The term ‘empowerment evaluation’ was originally proposed and described by Fetterman (2001; Fetterman & Wandersman, 2005) and has since been more widely adopted. It describes the use of evaluation concepts, techniques, and findings to foster improvement and self-determination. It is driven by a vision of the evaluator as an agent of social change (Miller & Campbell, 2006) and is designed to help people help themselves and improve their programs using a form of self-evaluation and reflection (American Evaluation Association - Collaborative Participatory and Empowerment Evaluation Topical Interest Group, 2005).

Using an empowerment evaluation approach, the goal is not only to evaluate the program in question but also to build the organisation’s capacity for future evaluation and continuous learning (Andrews, 2004). The evaluator is conceptualised as a ‘critical friend’ and the approach is based on a fundamental belief in people’s desire and capacity to understand their own situation and to create appropriate solutions, when provided with the necessary tools and conditions (Newell & Graham, 2010a).

Of course, empowerment evaluation, like other forms of evaluation, comes in different shapes and sizes. Miller’s study of 47 examples of published empowerment evaluation identified three different approaches:

- The **Socratic coaching approach**, where evaluators facilitate group processes. The group collectively decided on the evaluation aims, design, and procedures and collaboratively collected, analyzed, and reported evaluation data. The evaluator helps solve problems, provides training sessions, acts as a sounding board, and poses questions to the group that enhance its critical analysis of the program. The evaluator may help the group carry out activities but as a coequal to the other group members.

- The **structured guidance approach**, where the evaluator provides a set of steps, templates or worksheets and details of their implementation by way of manuals and/or training sessions. Program staff members and other stakeholders learn about evaluation by working through the provided templates.

- The **participatory evaluation approach**, where evaluators designed and executed most or all of the study, with program staff and stakeholders providing feedback on elements of the evaluation or participating in circumscribed ways, such as advising on the best way to recruit respondents, providing feedback on proposed measures, and helping with data collection. Evaluators do not provide training beyond that which is absorbed by participants in the process.

As we read in the previous section, empowering processes are typical of many participatory and collaborative approaches to evaluation (Miller & Campbell, 2006). As with other participatory approaches, stakeholders (potentially including clients) are directly involved in conducting the evaluation, supported by a facilitator, and there is a deep respect for people’s capacity to create knowledge about, and solutions to, their own experiences (American Evaluation Association - Collaborative Participatory and Empowerment Evaluation Topical Interest Group, 2005, p.2). However, what might distinguish empowerment evaluation is its broader impact. Empowerment evaluation, if properly carried out, should result in:
...actual shifts in power in terms of individuals’ engagement and participation in making decisions in which they previously were minimally involved or uninvolved and in an organization’s ability to garner resources and influence relevant policy (Miller & Campbell, 2006, p.298).

Empowerment evaluation is also not without its critics. Some raise concern about its reliance on self-study, its vague contingencies for practice, the rigor and propriety of the evaluations that may result from it, and the absence of rigorous evidence to show that it is indeed empowering and leads to empowered outcomes (Miller & Campbell, 2006).

The ‘most significant change’ technique

The ‘Most significant change’ (MSC) technique was developed by Rick Davies (through his work monitoring and evaluating rural development programs in Bangladesh) and has since been applied in many different sectors (including agriculture, education and health) and diverse cultural contexts. It is a form of participatory monitoring and evaluation which occurs throughout a program’s cycle and provides both formative and summative information to help people manage programs.

Essentially, MSC involves the collection of significant change (SC) stories at the field level, and the systematic selection of the most significant of these stories by panels of designated stakeholders or staff (Davies & Dart, 2005). There are two key questions which drive the MSC process namely:

Looking back over the last month, what do you think was the most significant change in [particular domain of change]?

From among all these significant changes, what do you think was the most significant change of all? (Davies & Dart, 2005)

Clear communication between each level involved in the program, and about the choices made in relation to the significant changes, is an essential component of the process.

Selecting approaches to evaluation

So, if there are all these types of evaluation models, and many more, which one is ‘right’ for your organisation and your program?

As flagged earlier, decisions about what type of evaluation is ‘right’ in particular contexts are, in part, ideological. They are also driven by the nature of the program and the purpose and audience for your evaluation. In some instances, where a program is funded by a government or other organisation, the expectations and/or requirements for the evaluation may be imposed, or there may be a sense that a particular approach may be more likely to be well received or lead to continued support:

What is at issue here... is not simply a disagreement over evaluation method but a fundamental issue of what is the proper purpose and social role of professional evaluation, who should control the profession, and to what ends (cf. Smith, 2004b). Whose values should control the shape of evaluation? (Smith, 2007, p.171)

Ideally, the evaluation approach should resonate with the nature and purpose of the program itself and should concord with the value system of those involved. For example, if the evaluation is of a health promotion program, experimental approaches which objectify participants may be considered inconsistent and participatory, empowering and collaborative approaches may be preferred (Green & South, 2006). For this reason, participatory approaches, and in particular empowerment approaches, are particularly significant when we consider evaluation that is conducted with and for children and young people.
Section 3: Why involve children and young people in evaluation?

Youth participation in community evaluation research is an approach in which young people are active participants in the stages of knowledge development, including defining the problem, gathering the information, and using the results. When young people define their own problems rather than discuss the ones given by adult authorities; when they design their own age-appropriate methods rather than uncritically accept adult ones; and when they develop knowledge for their own social action and community change rather than “knowledge for its own sake”—when they work in these ways... it can raise their consciousness and their spirit and move them to action (Checkoway & Richards-Schuster, 2003, p.22).

Checkoway and Richards-Schuster (2003) present five reasons for youth participation in community evaluation research. They argue it:

- is a legitimate way to develop knowledge for social action, since young people want to take action and improve communities, community agencies need young people’s perspectives;
- can enable young people to exercise their political rights, consistent with the United Nations Convention on the Rights of the Child;
- can allow young people to share in the democratisation of knowledge, potentially breaking monopolies on knowledge development and enable youth to have some of the information they require for competent citizenship
- can prepare young people for active participation in a democratic society, promoting and motivating civic engagement
- can strengthen the social development of young people, by increasing their individual involvement, their organizational development, and their ability to create community change.

Children and young people can participate in evaluation in different ways, each involving different patterns of involvement and different roles. For example, they can be subjects, consultants, partners or directors (Checkoway & Richards-Schuster, 2003; Checkoway & Richards-Schuster, 2004). These roles are well summarised in the Table 5.

We discuss issues related to the ethics of involving children and young people in evaluation in Section 5.
Table 5: Possible roles of children and young people in evaluation

<table>
<thead>
<tr>
<th>Goals of youth involvement</th>
<th>Youth as Subjects</th>
<th>Youth as Consultants</th>
<th>Youth as Partners</th>
<th>Youth as Directors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defining the questions</td>
<td>Adults define questions</td>
<td>Adults define questions</td>
<td>Adults often define questions with or without youth input</td>
<td>Youth define questions with or without adult input</td>
</tr>
<tr>
<td>Creating the instruments</td>
<td>Adults create instruments</td>
<td>Adults ask young people for feedback on their instruments</td>
<td>Adults and youth may jointly create instruments</td>
<td>Young people create instruments with or without adult input</td>
</tr>
<tr>
<td>Collecting the information</td>
<td>Adults collect information</td>
<td>Adults collect information</td>
<td>Youth may help adults collect information</td>
<td>Youth collect information, adults may assist</td>
</tr>
<tr>
<td>Analysing the information</td>
<td>Adults analyse information</td>
<td>Adults analyse information</td>
<td>Adults take lead in analysis, young people may assist</td>
<td>Youth take lead in analysis, adults may assist</td>
</tr>
<tr>
<td>Disseminating findings</td>
<td>Adults disseminate findings mostly to professional audiences</td>
<td>Adults disseminate findings mostly to professionals with or without your input</td>
<td>Adults take lead in dissemination, youth may assist</td>
<td>Youth take lead in dissemination, adults may assist. Findings may mobilise other youth or create community change</td>
</tr>
<tr>
<td>Roles of young people</td>
<td>Young people are subjects of study</td>
<td>Young people play limited roles as consultants</td>
<td>Young people assist adults in roles such as information collection and dissemination of findings</td>
<td>Young people initiate and take lead in all stages of the process. Adults may or may not assist</td>
</tr>
<tr>
<td>Roles of adults</td>
<td>Adults take lead in all stages of the process</td>
<td>Adults play most of the key roles</td>
<td>Adults initiate and implement the process, but enlist youth to assist them</td>
<td>Adults may or may not play supportive roles, but youth make the decisions</td>
</tr>
</tbody>
</table>
Section 4: Methodology and methods in program evaluation

In Section 2 of this paper we discussed a range of types of evaluation and their underpinning ideologies. Now we will turn our attention to matters of methodology (the philosophical assumptions underlying the evaluation approach) and method (the specific procedures that are applied when conducting the evaluation). Our intention is only to include a broad introduction to these ideas; enough to provide you with a grounding to inform further reading.

Paradigmatic considerations

In reaching a sound appreciation of evaluation processes it is essential to understand at least a little about paradigms — a philosophical term which refers to a view, a general perspective or set of assumptions about the world and how the world can be understood. Understanding the nature and influence of paradigms can help evaluators become more aware of their biases and assumptions and assist them to make methodological choices which are appropriate to the nature of the evaluation context (Clarke, 1999).

Typically two paradigms are discussed in evaluation and each holds different understandings of the nature of reality (what is known as ‘ontology’).

The quantitative paradigm is based on an understanding that reality is objective and can exist independently of human perception. This is referred to as ‘realist’ ontology. It assumes that there is a single reality and that social researchers can find regularities and relationships and discover the causes of social phenomena (a focus on causality). Quantitative methodologies favour a ‘hypothetico-deductive approach’, whereby a theory is put forward in the form of a hypothesis and this hypothesis is tested to confirm or disprove the theory. In evaluation, this usually takes the form of measuring relative attainment of predetermined goals (Clarke, 1999).

The qualitative paradigm, sometimes called naturalistic inquiry (Guba & Lincoln, 2000), is based on what is referred to as ‘relativist’ ontology — a belief that there is not a single reality that can be subjected to objective measurement. Rather, from this perspective, individuals and groups construct their own versions of reality (referred to as constructivism) and the evaluator or researcher’s task is not to identify which version is the ‘truth’, but rather to ensure different versions are accurately recorded and reported.
Qualitative methodologies favour an inductive approach whereby researchers aim to derive broad generalisations from observed data. In evaluation, this takes the form of direct observation of program activities and interviews with participants without specifically focusing on pre-determined goals. Whereas quantitative approaches focus on whether a program has reached its goals, qualitative approaches are more likely to focus on why and how these goals were reached (Clarke, 1999).

You are no doubt beginning to recognise how the various types of evaluation which we discussed in Section 2 might ‘fit’ within this paradigmatic distinction. The ‘randomised control trial’ approach is clearly predominantly informed by a realist ontology and a quantitative paradigm, whereas participatory approaches tend to be relativist or constructivist in nature, and predominantly qualitative.

While some philosophers and researchers argue that these two paradigms are logically incompatible, since they represent mutually exclusive ontological and epistemological positions, we have already flagged the growing emphasis on mixed-mode evaluations. For example, it is possible to design a quasi-experimental evaluation design, where there may well be an experimental group and a control group, but the methods employed within that broad design might be qualitative and participatory (for the experimental group at least). As Patton (cited in Clarke, 1999, p.62) has argued:

"...the evaluator needs to be situationally responsive and methodologically flexible. This entails moving between different paradigmatic stances as and when the need arises. Consequently, evaluators need to be well versed in both qualitative and quantitative techniques."

**Methods**

The methods which are typically used in evaluation are shared with research. It is beyond the scope of this document to describe these in great detail but a brief description of the more common methods is provided in this section (informed by Clarke, 1999).

**Questionnaires** are one of the most frequently used methods in evaluation and can take on many forms, from paper-based to online or verbally-delivered, (such as where the evaluator reads the questions and records responses over a phone). They can contain a wide range of types of questions and response forms (such as rating scales, attitudinal statements, or Likert scales responses). Questions can be *closed* (with a defined number of options to choose from) or *open* (allowing respondents to answer in their own words). Questionnaires are suitable for collecting large quantities of structured and standardised data. Important considerations are the format, presentation, length, sequence and language used, and its suitability to the stakeholders who will complete the questionnaire. Questionnaire data is typically recorded in a spreadsheet program and data are analysed using statistical processes.

**Interviews** are also widely used in evaluation and are an important approach in constructivist evaluation. Interviews can range along a continuum from structured (with the same question asked in the same way of all respondents) through semi-structured to unstructured (which is more like a free-flowing conversation). Interviews are often recorded (using a digital voice recorder) and transcribed either in full or selectively to produce an interview transcript. As with questionnaires, interview questions need to be very carefully designed and sequenced. Much also relies on the relationships and interpersonal dynamics which arise between the interviewer and interviewee.
Mostly, when we refer to interviews, we mean one-to-one interactions, however sometimes interviews occur in pairs (for example, with children).

**Focus groups** are a form of interview, but with a group of people, typically between six to twelve. The discussion is moderated by the evaluator, however the group context allows participants to interact with each other’s ideas, raise issues and seek clarifications. Of course there are implications to this (some may see them as disadvantages) in that some individuals may not feel free to express their true views or may feel pressured (consciously or subconsciously) to take a particular stance, and there isn’t the same degree of confidentiality. Advantageously, focus groups are an efficient means of collecting data and a way of gaining insights into group dynamics (such as within a workplace).

**Observation** is likely to be used by evaluators regardless of other methods chosen, since their interaction with any program or workplace will lead them to gain an insight into physical settings and interpersonal interactions. Systematic observation is also a data collection method in its own right. Again, this approach can be structured (for example, via a checklist or other template outlining what is being observed) or unstructured (where what is observed emerges from the context and action itself). Evaluators can find themselves in a number of roles when engaging in observation. They might be solely in the role of observer (sitting aside from what is happening and just observing) or they may be a ‘participant observer’, taking part in the activities and observing at the same time. Both approaches have strengths and weaknesses.

**Document analysis** involves close examination of primary and/or secondary documents associated with the program. Primary documents are those compiled directly by those involved in the program, such as manuals or training materials, minutes of meetings, diaries or memos. Secondary documents are those produced by individuals without first-hand experience, such as an externally written report or a funding brief. Such documents are often an invaluable source of information about a program’s goals and objectives.

### The concept of ‘baseline’ data

Many forms of evaluation, both quantitative and qualitative, use a design by which baseline information is collected before, or very soon after, a program begins. This data might be collected using any of the methods cited above, or other methods as appropriate. This baseline data provides a reference point to understand what might change as a result of implementing the program.

### Sampling

It is important to consider who will be involved in your evaluation, and how these participants will be selected. One component to these decisions is to identify which stakeholder groups have important perspectives to contribute. As previously indicated, stakeholders might include local practitioners or staff, clients or participants in the program (including children, young people and their families), other beneficiaries, representatives from other agencies or funding bodies, and sometimes the general public. Again, stakeholders vary, depending on the nature of the program.

However, it is generally not enough to simply decide to involve, for example, participants in the program. Normally, further decisions are required about which participants. In some instances, where a program is quite small, this may be *all* participants... but this is unusual.
Sampling is the process of selecting who will be invited to be involved in the evaluation. Broadly, a sample can be random or purposeful. We discussed random sampling approaches briefly before in relation to randomised control trials. Random sampling is required where the evaluator seeks to generalise from the study to a larger population (such is the case with opinion polls). Purposeful sampling does not allow generalisation but is used when the evaluator seeks more in-depth information.

Patton (2002a) identifies a wide range of different types of purposive sampling, each of which is useful in particular circumstances. Very briefly, we have summarised here some of the more commonly used approaches, relating them to the field of program evaluation:

- **Extreme or deviant case sampling** – cases that are rich in information because they are special in some way (e.g. outstanding success or non-success in a program).

- **Intensity sampling** – as above, but with less emphasis on ‘extreme’. Cases are selected as interesting and rich but not necessarily unusual (e.g. participants who might have made significant use of the program).

- **Maximum variation sampling** – aims at capturing central themes across a great number of cases and so involves identifying diverse characteristics of (for example) participants and sampling to include this variation (e.g. sampling from diverse geographical areas or diverse age groups).

- **Homogeneous samples** – used where the focus is on a particular subgroup (e.g. a focus on female single-parents within a broader group of parents). Focus groups often use homogeneous sampling.

- **Typical case sampling** – selected with the cooperation of key informants (such as program staff) who can help identify what might be a ‘typical’ case.

- **Stratified purposeful sampling** – where a typical case sampling approach is combined with another strategy to thus include above average, average and below average (e.g. including some ‘typical’ participants in a program, together with those for whom the program really worked well, and those who it didn’t work well for).

- **Critical case sampling** – those cases that can make a point dramatically or are otherwise particularly important (e.g. focusing on a particular staff member who is known to have encountered problems implementing a program).

- **Snowball or chain sampling** – having located one person or case, this person refers you to other people who are like them (e.g. in a circumstance where people attending a youth event are unknown and you speak to one young person who attended the event, and they refer you to others who also attended).

- **Criterion sampling** – where all cases that meet a predetermined criteria are sampled (for example, sampling all participants who remained involved with a program for more than 20 weeks).
- **Theoretical sampling** – where participants are selected on the basis that they have something relevant to say about a particular theory of focus (for example, if it is theorised that the program is particularly beneficial for males between ages 15-18).

- **Confirming and disconfirming cases** – this strategy is often used over time as theories emerge and the evaluator looks for particular cases which confirm or disconfirm the theory (e.g. a program was found to lead to high employment rates and the evaluator looks for participants who did not gain employment, to help explain why).

- **Opportunistic sampling** – Where the evaluator has to take advantage of whatever unfolds (e.g. in an evaluation of a media campaign, and data needs to be collected quickly, so young people who turn up to a youth event are approached).

- **Purposeful random sampling** – the sampling is still random, but from a purposeful subset, so it is not a representative random sample (e.g. choosing every second person on a list of a subset of participants in a program).

- **Convenience sampling** – doing what is fast and convenient. This is probably the most common approach and perhaps the least desirable since it is neither purposeful nor strategic.

Of course, an evaluation might employ a mixture of these approaches.

Another key question is the size of the sample. Quantitative evaluations have particular statistical formulas which can be applied to identify appropriate sample sizes. However, as with all types of evaluation, appropriate sample size will vary depending on what you want to know, the purpose of the evaluation, what is at stake, what will be useful, what will be perceived to have credibility, and what can be done with available time and resources (Amato, 2010). In-depth information from a small number of people can sometimes be more valuable than less depth from a large number of people. If you are not trying to generalise to a population (which, as we explained, qualitative research does not do) then it is less relevant to have very large samples.

An important concept to understand in relation to sampling in qualitative research and evaluation is the notion of ‘data saturation’. This is the point at which the evaluator knows that they are hearing the same things over and over from participants. At such a point, the sample might be considered to be adequate. Note that this is not to claim that if the study continued that fresh ideas wouldn’t emerge, but just that it is less likely.
Ensuring quality and credibility of data and findings

When designing an evaluation, consideration needs to be given to the quality and credibility of the data.

Traditional forms of enquiry have emphasised objectivity, internal and external validity and reliability (Patton, 2002a). Validity refers to the extent to which an evaluation actually measures what it sets out to measure. Internal validity relates to detecting change and being able to attribute this change to the intervention. External validity (or generalisability) concerns the extent to which the findings are applicable to other situations and contexts (Green & South, 2006). Where evaluation uses quantitative methods, specific statistical tests and processes can be applied and followed to enhance validity and reliability of findings.

Qualitative inquiry uses alternative criteria when ensuring quality and credibility of data and findings. Constructivist enquiry emphasises dependability (having a systematic process which is methodically followed) and authenticity (reflexive consciousness about one’s own perspective, appreciation for the perspectives of others, and fairness in depicting constructions in the values that underpin them) (Patton, 2002a).

Guba and Lincoln (2000) describe four criteria for ensuring quality and credibility in inquiry, which might be explained as follows:

- **Credibility**: Do the data sources (i.e. the participants) find the inquirer’s analysis, formulation and interpretations to be credible and believable?
- **Transferability**: While discounting the idea of generalisability (i.e. that results can directly be applied to other contexts) qualitative inquiry asks: Is enough ‘thick description’ of the context and circumstances provided so that some reasoned judgements can be made about the degree to which the findings are relevant/transferable to their own context?
- **Dependability**: Are the findings ‘stable’ after discounting both conscious and unpredictable changes?
- **Confirmability**: The idea of objectivity is dismissed and it is not the inquirer’s objectivity which is of interest but whether the data can be confirmed.

Guba and Lincoln (2000) describe a range of strategies which can be used to enhance credibility, transferability and confirmability. These include strategies for credibility such as prolonged engagement at a site, persistent observation, peer debriefing, triangulation (explained below), referential adequacy of materials and member checks (explained below). Transferability is enhanced through theoretical/ purposive sampling and thick description. Dependability is enhanced through use of multiple methods, stepwise replication and dependability audits. Confirmability is enhanced through triangulation, practicing reflexivity, and confirmability audits (Guba & Lincoln, 2000).
Triangulation, the practice of seeking more than one reference point or multiple perspectives, is a key feature of qualitative evaluation methods, enabling greater accuracy of understanding. Triangulation might entail using more than one of the methods above or asking similar questions of different groups of stakeholders. Triangulation can also be achieved when more than one evaluator or researcher investigates the same situation, ensuring that different viewpoints are represented. Another approach is to collect data at multiple points of time. Triangulation allows the evaluator to have greater confidence in their findings.

Member checking is another strategy which should be explained in a little more detail. This is a strategy where data, and interpretations of this data, are continually checked with members of the various stakeholder groups. It is done continuously throughout the evaluation and again at the end when the full report is compiled (Guba & Lincoln, 2000).

Participatory approaches to evaluation, which we discussed earlier, require particular approaches to judging quality. As with other qualitative approaches, the idea of objectivity is usually rejected in favour of a holistic approach that incorporates the diverse perspectives, values, agendas and interpretations of participants, stakeholders and evaluation professionals (Lennie, 2005). Criteria such as trustworthiness, including credibility, transferability, dependability and confirmability are particularly emphasised. The value and ‘useability’ of the findings to the participants themselves is also of critical importance.

The role of the evaluator

Thoughout this paper we have discussed the many and varied approaches to program evaluation and already you would have ascertained that the roles of evaluators vary considerably between these models. In some approaches there isn’t a separate ‘evaluator’ role at all, but rather those participating in the program share in the role of evaluating it.

Nonetheless, it is useful to briefly consider the many roles which evaluators can, and do, play. Table 6 is drawn from Skolits, Morrow and Burr (2009).

Table 6: The roles of evaluators (from Skolits, Morrow & Burr, 2009).

<table>
<thead>
<tr>
<th>EVALUATION ACTIVITY</th>
<th>EVALUATION PHASE</th>
<th>EVALUATOR’S PRIMARY ROLE</th>
<th>POTENTIAL SECONDARY ROLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation management</td>
<td>All phases</td>
<td><strong>MANAGER</strong> Address evaluation administration &amp; role coordination</td>
<td>• Diplomat • All roles</td>
</tr>
<tr>
<td>Initial client contact</td>
<td>Pre-evaluation Phase</td>
<td><strong>DETECTIVE</strong> Determine evaluation need &amp; alignment with evaluator skills, competencies &amp; interests</td>
<td>• Manager • Diplomat • Use advocate</td>
</tr>
<tr>
<td>Evaluation contracting</td>
<td></td>
<td><strong>NEGOTIATOR</strong> Agree an evaluation contract between evaluator &amp; client</td>
<td>• Manager • Diplomat • Other roles</td>
</tr>
<tr>
<td>Evaluation planning</td>
<td>Active evaluation phase</td>
<td><strong>DESIGNER</strong> Develop a realistic &amp; responsive evaluation design to address the client’s need</td>
<td>• Manager • Diplomat • Reporter • Other roles</td>
</tr>
<tr>
<td>Initial implementation</td>
<td></td>
<td><strong>DIPLOMAT</strong> Establish trust &amp; rapport with stakeholders</td>
<td>• Manager • Diplomat • Use advocate</td>
</tr>
<tr>
<td>Data collection &amp; analysis</td>
<td></td>
<td><strong>RESEARCHER</strong> Collect &amp; analyse reliable &amp; valid data</td>
<td>• Manager • Diplomat</td>
</tr>
<tr>
<td>Evaluation judgment</td>
<td></td>
<td><strong>JUDGE</strong> Make evidence-based judgement of worth &amp;/or opportunities for improvement</td>
<td>• Manager • Diplomat • Researcher</td>
</tr>
<tr>
<td>Evaluation reporting</td>
<td></td>
<td><strong>REPORTER</strong> Share evaluation results with appropriate stakeholders</td>
<td>• Manager • Diplomat • Researcher</td>
</tr>
<tr>
<td>Evaluation use advocacy</td>
<td>Post-evaluation Phase</td>
<td><strong>USE ADVOCATE</strong> Promote &amp; support improvement &amp; change</td>
<td>• Manager • Diplomat • Other roles</td>
</tr>
<tr>
<td>Evaluation reflection</td>
<td></td>
<td><strong>LEARNER</strong> Reflect on the evaluation to improve future practice</td>
<td>• All roles</td>
</tr>
</tbody>
</table>
Bringing it all together: Planning for program evaluation

So having considered all these aspects of program evaluation methodology and methods, what is actually involved in planning an evaluation? The more complex the program in question, the greater the need for good planning and focused attention to the methods employed.

Stufflebeam (2000a; 2003) has proposed a model for guiding formative and summative evaluations. The CIPP model has four core parts (with its corresponding acronym), and each has a key respective question. The fourth is divided into sub-parts as follows:

- Context - What needs to be done?
- Input - How should it be done?
- Process - Is it being done? and
- Product - Did it succeed?
  - Impact - Were the right beneficiaries reached?

Stufflebeam produces a checklist to assist groups to plan, carry out, institutionalise, and/or disseminate evaluation, and to help evaluators review and assess a program’s history and issue a summative report on its merit, worth, probity, and significance, and the lessons learned (Stufflebeam, 2007).

Another useful framework to assist you to get started on planning an evaluation has been developed by researchers at the Centre for Children and Young People at Southern Cross University. Newell and Graham (2010b) propose a cyclical model to depict the key stages and questions that inform program evaluation design and planning. This is provided in Figure 2.

Figure 2: The cyclic model of program evaluation

![Cyclic Model of Program Evaluation](image-url)
Section 5: Ethics in Program Evaluation

Although we arrive at this topic late in the Background Briefing, this is not to imply that it is not of foundational and central importance. Evaluation, as with all forms of research, must place ethics first, foremost and central to practice.

The principle that underpins ethical research is the view that research is not just a matter of collecting information, but is concerned with the dignity, rights, safety and well-being of those who take part in research (Stuart & Barnes, 2005).

There are a wide range of ethical concerns which can arise in evaluation. Turner’s (2003) survey of 132 evaluators revealed the following as issues which had been encountered:

- Requests to use information gathered for one purpose (e.g. program improvement) for a different purpose (e.g. accountability)
- Unilateral changes to terms of reference midstream or at time of reporting an evaluation and dealing with the implications for quality and relevance of data collected
- Issues related to dealing with different types of evaluation subjects or participants
  - Working with indigenous people
  - Research with children
  - Sensitive topics such as sexual victimisation
  - Feeding back results to participants
  - Informed consent
  - Privacy and confidentiality
  - Interview subjects disclosing confidential or inappropriate information in interviews

Other
- Surfacing issues of incompetence or poor performance among program staff
- Whether or not to use ethics committees

In this section we consider, although by necessity briefly, just some of the key ethical issues which must be taken into account, namely: weighing up benefits and risks; access to stakeholders; informed consent; privacy, confidentiality and anonymity; cultural considerations; payment to participants; child protection and dissemination and reporting.
Benefits of evaluation are often much easier to identify, since most times the evaluation is aimed at improving services and programs. However, there are instances where an evaluation of a program may result in the program being withdrawn or discontinued. Further, it is important to be clear that sometimes improvements don’t actually result for those individuals participating in the evaluation. Rather, the evaluation might only lead to future improvements in service provision, too far into the future for those individuals involved.

Access to stakeholders

One of the key ethical issues is who should be involved and how you will gain access to these groups of stakeholders. When wanting to involve children, young people or families particular issues arise in relation to access. Matters of confidentiality often need to be considered where individuals have been accessing a service or program. Evaluators need to heed who has legitimate access to records related to these individuals, and who is in a position to invite their participation. This may not be the external evaluators. Another key question is whether the evaluation is seeking to include individuals who are not using a service, how they would they be chosen and what benefits there are for them.

Sometimes there are gatekeepers who need to be consulted and their permissions and guidance gained. In relation to schools, gatekeepers exist at many different levels. For example, in Australia, State Departments of Education generally have official policies, procedures and guidelines that must be followed across all schools, which can involve gaining approval from regional education authorities, as well as school principals.

However, evaluation isn’t always dealt with in the same way as research in these guidelines. In many instances, quality assurance and evaluative processes are viewed as a normal part of good professional practice, and may not require all the same levels of approval as (say) research.

Weighing up benefits and risks

The concept of risk or harm needs to be carefully considered before embarking on any evaluation, particularly that involving children and young people. The concept of ‘harm’ is difficult to define, as the issues will vary depending on the nature of the program being evaluated. However, it is important to recognise that harm can be psychological as well as physical.

Whilst physical harm needs little in the way of explanation, psychological harm can be more complex. Examples of psychological harm might be if participants are embarrassed or humiliated or placed under undue stress during the course of a study. Less obviously, questionnaires that ask people to reveal their attitudes or beliefs, or personal information about income, religious or political affiliations can cause anxiety for some respondents. You need to be sensitive to the fact that some topics can have unexpected consequences (e.g. talking about a stressful pregnancy, or about reasons for not being able to access a service, may lead respondents to relive upsetting events in their lives) (Stuart & Barnes, 2005, p.7-8).
Informed consent

Informed consent goes beyond the initial process of ascertaining broad interest or willingness to participate. Participants need to have the evaluation purpose and process explained to them fully, and be invited to join in on the basis of this knowledge. They should have their rights explained and be told that if they do not want to participate they will not be disadvantaged.

They also need to be given permission to change their mind at any time and for there to be no consequences or retribution for not participating. It is important not to assume that people understand what is involved in an evaluation and that individuals have the opportunity to ask questions or to have explanations repeated (Stuart & Barnes, 2005).

In most cases, participants should sign an informed consent form saying that they understand what the project involves and their rights. Since this form includes the participant’s name it should be kept separately from their data in a locked filing cabinet (Lienert, 2002). The information sheet should include a description of the evaluation; what it means for the participant; how long participants will spend answering questions/ filling in a questionnaire; information on confidentiality; whether taking part is compulsory and details of those to contact to ask more questions (Stuart & Barnes, 2005).

When you are involving children and young people, the principle of informed consent still applies, although the means of informing and gaining consent may vary, depending on the age of the child. Simple and age appropriate language is essential and the process can be done verbally rather than in writing if necessary. Of course, if you are involving children then you will also require the consent of a parent or guardian and may need to consider parents who have low literacy levels and that they only agree to take part, or for their children to take part, after fully understanding what is involved.

Privacy, confidentiality and anonymity

If people give details about themselves, these details should be confidential, and kept in a safe place where no one except the evaluator can access them. No information about individuals should be stored with the data that could identify them. Case notes should have some details changed so that no one will know who the data is about.

The way in which confidentiality is dealt with does vary according to the methodology being used. For example, in focus groups, where people are being asked to express their opinions to people they may or may not know, it will be important to undertake some discussion with the group about what confidentiality means in these circumstances (Stuart & Barnes, 2005).

A related aspect is that of anonymity when reporting on the evaluation. It is important that any participant’s opinions are not attributed to identifiable individuals. Some evaluators use only a description of the respondent (e.g. girl aged 6) while others use pseudonyms (made-up names). Children themselves can be involved in choosing their own pseudonym as part of the process of discussing anonymity and confidentiality.
Cultural considerations

Schwandt (2007) argues that evaluators must also heed matters of culture – with culture defined in its broadest sense:

Both our institutions (the economy, politics, the family, education, etc.) and social practices (teaching, evaluation, administration, nursing, etc.) embody and specify cultural values in terms of what is right and wrong, good and bad—they mediate our relations with others and the world.

Schwandt (2007) describes how, when encountering situations where one’s norms and values are at odds with another’s, the evaluator will commonly respond in one of two ways: either maintaining their own ethical norms while politely respecting the norms of others; or claiming that one can adapt to the norms outside one’s culture. The latter, Schwandt argues, ‘...while seemingly a noble ideal, is based on the faulty assumption that one can literally be one other than oneself (i.e., walk in another’s shoes)’ (Schwandt, 2007, p.402). Rather, Schwandt argues for a model of cultural dialogue, where participants learn from one another through discussions and disagreements, about ethics and norms.

Payment to participants

Whether or not participants in an evaluation should be paid for their time is a difficult question and very much dependent on the nature of the program and the participants. It is sometimes thought acceptable to offer some recompense but payment must not be used to induce people to participate or to influence the results of the evaluation. In some instances a ‘thank you’ voucher or gift, provided after participation rather than before, can overcome some of these issues.

Child Protection

As with any activity which involves working with families, children and young people, those involved in evaluations need to be aware of child protection concerns. In some cases evaluators will need to gain the required clearances to work with children. Further, they need to be aware of their obligations in terms of reporting any concerns they may have. In research and evaluation activities, child protection raises particular issues in that it is not possible to guarantee a child or young person that you would treat any such disclosures as confidential.
Section 6: Future directions

Public spending on children, young people and families has grown considerably in recent years. Interventions and programs that aim to improve children’s lives are increasingly required to demonstrate they are effective. Yet government and non-government organisations continue to grapple with why key indicators (such as engagement with school, substance abuse, antisocial behavior, mental health etc) often suggest deterioration rather than improvement in many children’s circumstances.

Recognising the turbulent contexts in which such programs operate, and designing appropriate evaluation approaches that can take account of these complex influences on program implementation, is one of the most significant challenges for evaluators in the 21st Century.

It is not insignificant that Michael Patton, one of the key theorists in the area of program evaluation, has recognised such issues in his most recent publication (Patton, 2011). As he explains:

*Change involves a belief in the possible, even the ‘impossible’. Moreover, social innovators don’t follow a linear pathway of change; there are ups and downs, roller-coaster rides along cascades of dynamic interactionism, unexpected and unanticipated divergences, tipping points and critical mass momentum shifts. Indeed, things often get worse before they get better as systems change creates resistance to and pushback against the new.*

*Traditional evaluation approaches are not well suited for such turbulence. Traditional evaluation aims to control and predict, to bring order to chaos. Developmental evaluation accepts such turbulence as the way the world of social innovation unfolds in the face of complexity. Developmental evaluation adapts to the realities of complex nonlinear dynamics rather than trying to glimpse order and certainty on a disorderly and uncertain world (Patton, 2011, p.5).*

These new understandings of both programs and their evaluation are likely to bring significant changes to the theory and practice of evaluation over coming years.

So what does all this mean for practitioners who are seeking to develop the most appropriate measurement tools and ensure interventions translate into meaningful outcomes for children and young people?

Most importantly, it heralds the need to continually critique the forms of evaluation that we use in our workplaces and community contexts. We need to critically assess the ideologies, criteria, methods, findings, conclusions, recommendations and interpretations of evaluations. We need to be particularly conscious of the context and circumstances within which the evaluation takes place, and what stakeholders want and need from the evaluation (Stufflebeam, 2000c).

Recognising that program evaluation plays a critical role in effectively meeting the needs of children, young people and their families and is an essential first step in the ongoing commitment to improving our professional practice. We hope this Background Briefing provides a useful resources for you in continuing this journey.
References


Centre for Community Child Health - Royal Children’s Hospital, & Broadmeadows Early Years Partnership (2007). *Communities for Children in Broadmeadows: Optimising the health, development, learning and wellbeing of children living or participating in the Broadmeadows community*. Melbourne: Royal Children’s Hospital.


Evaluation, 29(4), 427-442.


About the Centre for Children and Young People

The Centre for Children and Young People (CCYP) was established at Southern Cross University in 2004. The CCYP works collaboratively with organisations, particularly in regional and rural areas, to enhance policy and practice related to the well-being of children and young people.

The Centre has three priority areas: Research, Education and Advocacy.

For more information about the CCYP, visit ccyp.scu.edu.au

About the Course

The Graduate Certificate, Graduate Diploma and Master of Childhood and Youth Studies are awards which have been developed collaboratively by the Centre for Children and Young People and the School of Education at Southern Cross University, Australia. The awards meet a recognised need, expressed by a range of professionals, for contemporary knowledge and skills to assist them to work more effectively with children, young people and their families.

The course seeks to be an innovative, professionally relevant, practical and interdisciplinary qualification for people working, or intending to work, with children, young people and their families. Applicants can enrol in any one of the awards or complete individual units as professional development.

Units are delivered externally so that students can successfully study at a distance. Each unit has authentic and professionally relevant assessment and the five core units involve optional but highly recommended summer/winter intensive workshops of 2 days duration. Students who are unable to attend are able to engage with workshop content and processes live online or via recorded formats.

The course incorporates innovative and appropriate use of technology to support students’ learning, opportunities for regular engagement with tutors and fellow students and (where appropriate) multimedia elements.

The course is underpinned by a deep respect and regard for children and young people and for their views and perspectives. It also incorporates an understanding that children and young people can benefit immensely from positive relationships with adults – parents, teachers and the myriad professionals with whom they may engage over the course of their childhood. The course embraces multidisciplinary perspectives in the belief this can enhance service provision and lead to improved outcomes for children and young people.

For more information about these awards, visit www.scu.edu.au/childhoodstudies