



ourcommunity.com.au

Measuring What Matters

An introduction to project
evaluation for not-for-profits

www.ourcommunity.com.au





Measuring What Matters

An introduction to project evaluation for not-for-profits

Measuring What Matters: An introduction to project evaluation for not-for-profits

Published by Our Community Pty Ltd

© Our Community Pty Ltd

This publication is copyright. Apart from any fair use as permitted under the Copyright Act 1968, no part may be produced by any process without permission from the publisher

Requests and inquiries concerning reproduction should be addressed to:

Our Community Pty Ltd
PO Box 354
North Melbourne 3051
Victoria, Australia

Or email service@ourcommunity.com.au

Please note:

While all care has been taken in the preparation of this material, no responsibility is accepted by the author(s) or Our Community, or its staff, or its partners, for any errors, omissions or inaccuracies. The material provided in this guide has been prepared to provide general information only. It is not intended to be relied upon or be a substitute for legal or other professional advice. No responsibility can be accepted by the author(s) or Our Community or our partners for any known or unknown consequences that may result from reliance on any information provided in this publication.

ISBN 978-1-876976-61-3

Published March 2017

Cover photo by iStockphoto/DNY59

Contents

Terminology	7
Evaluation principles	8
Evaluation frameworks	12
What element of the project are we evaluating, exactly?	16
Process evaluation	18
Output evaluation	19
Outcome evaluation	21
Impact evaluation	22
Gathering the data	27
Dealing with complexities	32
The evaluation cycle	34
Sample evaluation reports	37
Further reading	39

Introduction

If something is worth doing, it's worth evaluating. A program that is properly evaluated gives twice: first, in the impact it has, and second, in making it possible for you – and others – to do better next time.

You have a dream – that the world is going to adopt your philosophy, or children are going to drink more milk, or people will stop eating animal products altogether, or people with cancer are going to get more of a particular type of help, or the Goodtown Pumas are going to win the final – and that dream is not going to happen unless you make it happen.

To make the world different needs resources – time, effort, equipment, petrol, attention, pencils, webpages, all sorts of things. You may have to buy them, or you may have to ask people for them, or you may have to make them happen yourself. However you assemble your resources, there will never be enough of them, because that is the nature of this universe. Other people have other priorities. Any resources that are spent on one project are not available for another. All of us have to make choices about where to allocate what we have at our disposal.

Before you make that choice, ideally you estimate what the outcome would be if you took this path or that, compare the different methods, and then pick the one most likely to be most productive. After you've done the project you look at what the outcomes were, you ask whether they were what you expected, and you hope that this knowledge will help you to make the next choice when it rolls around in an hour or a day or a year.

That's evaluation.

It may be that other people also want to know how well you did. Your grant funder or a donor may want reassurance that their money is going to the right place. The government may want to confirm that you're meeting the terms of your service contract. They have to make choices, too, about the distribution of their own resources, which are also not limitless. If they'd funded some other group – and there are always other groups – would they have done better?

The answer is seldom straightforward. There are invariably variables – possible explanations at different levels on different scales. Would the Pumas have done better against Badtown with a different coach or a more expensive energy drink, or would none of that have made a difference once the wool price collapsed, the local bank closed, and the Goodtown economy went off the boil? The world is a big blooming buzzing confusion, and it's hard to keep track of your own part when everything else is swapping places around you.

Nonetheless, you have to make sense of it all – enough, at least, to decide what to do next. To make good decisions you have to gather the data, and group it, and analyse it, and impose value on it. What are you doing right? Where could you do better? Of the things you could do better, which is more important?

For people, for programs, for organisations, data is a torch in the fog.

EVALUATIONS THAT MAKE A DIFFERENCE

A top-notch project evaluation is not an end in its own right. Ultimately, what makes project evaluation worthwhile is making a difference in people's lives.

In 2014–16, a project called Evaluations That Make a Difference set out to collect stories of evaluations that did just that.

Women giving birth in remote areas of Papua New Guinea; small and medium businesses in Sri Lanka; community radio audiences in Nepal – these are among the beneficiaries whose stories are told at the website Evaluations That Make a Difference. Read more about the project and the stories at <https://evaluationstories.wordpress.com>.

Terminology

In the world of evaluation, terminology can be confusing. Different evaluation models have different names for each stage of the process. In this book, we've adopted the most common definitions, as follows:

Inputs: The resources used to complete a project, such as money and volunteer hours. How much money did we raise in the campaign to repair the pavilion roof? How many trees were donated for the tree-planting day? How many people volunteered to take part?

Mapping or counting the resources used across the operation – your inputs – can be laborious, but it's generally not conceptually taxing.

Activities: the things that you do with the inputs – run workshops, rescue cats, counsel troubled teenagers, plant trees. Some project evaluation models slot activities under “inputs”, and some regard them as “outputs”, but we've put them in their own category.

Outputs: the immediately discernible products, goods, services or achievements brought about by the inputs and activities – 20 cats rescued, 2000 trees planted, three workshops conducted, 65 teenagers counselled.

Outcomes: the medium-term effects that those products, goods, services or achievements have when fed out to individuals and into the social environment; for example, a 30% reduction in the incarceration rate of teenagers in Troubledtown. Remember that outcomes might be intended or unintended. An outcome of the rescue of 20 cats might be a 500% increase in the rodent population.

Impacts: the long-term developments that can be attributed to what you have done. How much has changed because of the work you put in? In what way is the world a better (or different) place? For example, the impact of the counselling of 65 teenagers might be increased social cohesion in Troubledtown 12 months later.

You'll have to pick your time horizon; the week after; the year after; five years on, the next generation? The further you go, the harder it becomes to see your faded signature on the benefits that arise.

Avoiding confusion

Some evaluation materials conflate output and outcome, and some blend outcome and impact. Some, indeed, use the term “impact” for medium-term results and “outcome” to describe the final population-wide effects (the reverse of our terminology).

To avoid confusion, you should define how you use these terms in any discussion or publication of your evaluation.

Evaluation principles

Mission-centred evaluation

You know that “aims” clause in your constitution, setting out all the things you want to do? Have you done all those things? How do you know?

Your entire purpose as an organisation is to achieve your mission. Everything you do – everything – should be directed to that end.

You must, at some point, be able to say whether you’re doing what you said you’d do. And you will need to present some evidence in support of your claims. Evaluation is the process of collecting this evidence.

In part, you will need evidence of your achievements to satisfy external stakeholders – grantmakers, government, donors, volunteers, business partners, and the general public. All of these groups demand evidence-based decision-making. Their expectations are high, and are often embodied in funding contracts.

More importantly, though, you will want to evaluate your success to assure yourself that you are in fact doing as much good as you think you are. If you really want something to be true, it is easy to deceive yourself into believing that it is true, and only unbiased evaluation can ensure that you are not wasting your time and the community’s resources on vanity projects.

Evaluation comes first

If you don’t build evaluation in from the beginning, you’ll have a hell of a time trying to bolt it on later.

Evaluation must be featured in your project design process from the very earliest stages. Budgeting and timetabling implications must be assessed and included in the design.

You should try to establish the parameters of your evaluation precisely from the beginning. Having said that, it is not always possible to establish in advance what the central features of a project will turn out to be, and no evaluation can cover every possible set of events. Include some flexible and open-ended opportunities to allow unforeseen developments to be caught up by your examination.

Once you’ve decided what you want to measure, the first question is what readings you need to take to establish a baseline. The figures

that you come up with afterwards are meaningless unless they can be compared with the pre-existing situation.

Resourcing your evaluation

Evaluation needs resources, including time, money, patience and – importantly – commitment to acting on the results. If you read an evaluation report, think “that’s interesting”, then lock it in a cupboard, the evaluation has been a waste of time.

Many elements of evaluation are expensive. Surveys and focus groups are expensive. Even if you do no more than collect client satisfaction forms at the end of an afternoon, someone is going to have to design them, distribute them, distribute them again when people forget to fill them in or ignore your requests to do so, collate the responses, and analyse the data. You may need new software, new hardware, or external expertise.

All this costs money that you then don’t have available to spend on the pointy end of the project, and you may begrudge the expense. You have to strike a balance. It is never appropriate to have no evaluation at all, and it is always undesirable to spend more on it than necessary.

The worst possible approach, however, is to include “evaluation” in a project design without allocating sufficient money to fund it. In such a case you will end up with neither the data nor the money. You must budget adequately for whatever evaluation you decide on.

Some grants schemes require, at the point of application, an outline of your projected evaluation, but most do not. You may have difficulty persuading your funders that this expenditure is justified, and you must be prepared to make the case for it (whether or not they raise it). If there is no other space on the application form, include evaluation in the project outline and budget, and explain that the data will be a valuable element of the total project.

Thankfully, more and more funders are coming to see evaluation as not only a legitimate but a desirable expense. If funders are really engaged in the quest for improvement rather than simply getting the money out the door, they will appreciate your efforts.

Evaluation needs commitment

Will you do what the evaluation tells you to do?

Evaluation takes time, money, and patience. No unnecessary evaluation, and no unnecessary element in any evaluation, should be permitted.

“Unnecessary” in this context means “we don’t care about this enough to modify our policy”. Unless the data you are collecting could in some circumstances lead you to change your practices, your procedures, or your policies, then there is no point in collecting it.

Asking this question (“Do we care enough about this that we might modify our policy?”) can be useful in cutting back on your information demands. If you find that a survey form, for example, goes directly into the filing cabinet to gather cobwebs, then something is badly wrong, and unless you are prepared to put some time into analysing it then you should strike it from your repertoire and save everybody down the line valuable time.

VIRTUAL INFANT PARENTING: THE COST OF FAILING TO EVALUATE

The results of some evaluations seem so predictable that it is tempting to see the research as a waste of time. “Literacy linked to higher employment”, “Credit card debt spiral causes mortgage stress” – most of us would have foreseen those results. It can be tempting to subscribe to the commonsense school of thought and skip the evaluation.

A long-term study of teenage girls in Western Australia, however, is a good example of why evidence-based evaluation is so critically important.

Starting in 2003, researchers set out to evaluate the effectiveness of a program intended to reduce the rate of teenage pregnancies. The Virtual Infant Parenting (VIP) program required schoolgirls aged 13–15 to care for a “robot baby”, a life-like simulator that replicated the sleeping and feeding patterns of a six-week-old infant. The “baby” displayed infant behaviours such as different forms of crying, which were programmed to occur with a similar intensity, frequency and duration as those observed in real infants.

The researchers hypothesised that the program would reduce the rate of teen pregnancies among participants. To test their hypothesis, they gathered data on the number of those girls who actually became pregnant by the age of 20. A control group of schoolgirls did not take part in the VIP program, and their pregnancy rate was also measured.

The researchers found that 17 per cent of girls who did the VIP program had become pregnant by the age of 20, compared to 11 per cent of those who did not take part in the program.

A commitment to following where evaluation leads will in many cases involve arranging to do better next time. This involves documenting the learnings from the evaluation, storing them until the opportunity recurs, and having filing and indexing systems that enable you to recover and apply them.

Every organisation has rules governing how practices, procedures, and policies are altered – some flexible, some not. These governance procedures must be held up against your evaluation outline to ensure that the results of any evaluation will be fed into the decision-making matrix and that there are no serious obstacles to its impartial consideration.

“The results were unfortunately not what we were hoping for,” lead investigator Dr Sally Brinkman told the ABC in 2016. “The aim of the program was to prevent teenage pregnancy. We can definitely say that it didn’t do that.”

Researchers also found 53.8 per cent of the pregnancies among girls who participated in the robot program were terminated, compared to 60.1 per cent of those who did not.

Dr Brinkman said the study did not explore why the program had failed, but anecdotally it seemed the girls enjoyed the process of caring for a baby. She recommended schools stop using such programs.

Outputs, outcomes and impacts: evaluate them all

If we evaluate the success of the program in terms of outputs – e.g. the number of schoolgirls who undergo the VIP program – then we might regard “1000 girls educated” as a more successful program than “500 girls educated”.

However, if we evaluate the success of the program in terms of outcomes – its effect on the rate of teen pregnancy – then we get a very different result (more teen pregnancies).

Now let’s look back to our earlier definition of impacts: “the longer-term effect of your activity. How much has changed because of the work you put in? In what way is the world a better (or different) place?”

The impact of our activity might consist of the lower education levels and higher rates of socioeconomic disadvantage associated with teen parenting. Measuring and evaluating impacts was beyond the scope of the WA study. However, known educational and socioeconomic indicators could be used as proxy measures.

Evaluation frameworks

When you think of evaluating a project, perhaps the first thing you think of is gathering data. That's great, but data on its own is about as useful as ice-cubes in a warm bath. Data needs a reason for being, and it all comes back to your mission and vision: what is the change that your organisation wants to see? What are you going to do to bring about that change? And what makes you think that your intervention will work?

You can survey farmers and count chickens until the cows come home, but you need to be pretty sure that counting the chickens will be a useful indicator of the eggs that will be hatched if you want to increase the number of omelettes in the world.

This is where evaluation frameworks come in.

Logic models

One of the most widely used frameworks for evaluating social change projects is the logic model. A logic model explains how particular activities lead to particular outcomes – if we do this, it will (we believe) result in that.

A logic model is a useful way for you (and your funders and stakeholders) to test your assumptions about a project (or about anything, really). It provides a way for you to plot a causal chain from what you are proposing to do, all the way to your eventual goals.

The importance of precise goals

If you don't know exactly where you're going, any road will get you there.

In some ways, your evaluation planning doubles as a reality test for your project. If you can't design a way to evaluate it, it could be that your goals are simply too vague and unfocused – for example, "This project will improve health."

If you are having difficulties designing your evaluation, go back and ask yourself whether you need to tighten the project itself. Let's say your organisation's vision is a state free from tobacco-related diseases. Perhaps, when you say, "This project will improve health", what you really mean is, "This anti-smoking campaign aimed at migrant schoolchildren will reduce the rate of smoking-related disease among this particular population when they grow up."

Starting at the end point (you want a state free from tobacco-related

disease), you work back to find the interventions that you hope will lead to that outcome. That's your logic model. Here's how it might look:

You set up an anti-smoking campaign.

And as a result:

Campaign materials reach migrant schoolchildren.

And as a result:

Knowledge of smoking-related conditions among migrant schoolchildren increases.

And as a result:

Rates of smoking among migrant schoolchildren decrease.

And as a result:

Rates of smoking-related disease among these children when they grow up decrease.

And as a result:

Rates of smoking-related disease in the state decrease.

Before you make a decision on whether to run such a campaign, you need to be satisfied that all the links in the chain are sound.

The link between less smoking and less disease has been exhaustively documented; the link between greater knowledge and lower smoking rates in this population may require some literature citations or references to earlier campaign data; the link between getting the material and reading the material might need testing in advance; and the link between the campaign and the material reaching the target group would need to be spelled out in the plan.

Each link in the chain should be capable of being evaluated, and this is where data comes in. At the end of the anti-smoking project you want to be able to prove that

- The material did reach the target group (through survey data)
- The material did increase health knowledge (through in-depth focus group data)
- Smoking rates did diminish (through survey and sales data).

A thorough evaluation would also include a further survey years later to evaluate whether any observed improvement in smoking rates was sustained. What we actually want to see, of course, is less disease (that's the outcome you're chasing), and the link between lower smoking rates

and lower disease incidence cannot in this instance be demonstrated until several decades later.

Few projects can provide meaningful evaluation over such timeframes. In this instance, the weight of the scientific evidence that smoking causes disease makes smoking rates a very reliable proxy measure, but such robust correlations are not always available.

THE FIVE WHYS

The Five Whys is a Japanese management tool designed to push you to think more deeply about what you're doing. It involves behaving like a particularly irritating six-year-old – but it works. It may also assist you with differentiating between the successive levels of achievement you want to document.

The Five Whys exercise is particularly useful in helping you to arrive at your true desired outcome.

Why do you want the money?

So we can run a camp for local children at risk.

Why?

So they can get a chance to be involved in group experiences they enjoy.

Why?

So they can learn about co-operation and gain self-esteem and see new possibilities in life.

Why?

So they can build resilience strategies and cope with their bloody difficulties better!

Why?

So they have better mental health as adolescents and adults.

There, was that so hard?

What do you need to know?

A thorough evaluation will relate closely to your logic model. It will enable you to check at each stage that your expectations have been met. You can also work backwards from your goals, seeking measures that would indicate that they had been met.

How will you get the answers?

Because the range of not-for-profit goals is effectively infinite, it is impossible to list every method of evaluation you might find useful in your particular situation. You may need to employ social surveys, statistical overviews, or satellite photography. Your choice will be guided by cost, by relevance, by estimates of effectiveness, and by the constraints placed on your enterprise by considerations of ethics, privacy, and public relations. For more on the most common evaluation methods, see page 29.

What element of the project are we evaluating, exactly?

Many things feed into the success of a project, not least the capacity and performance of the organisation undertaking the work. It can be conceptually useful to consider all these different elements of a project separately.

EVALUATION: BUSINESS AS USUAL

A dog is for life, not just for Christmas. In the same way, evaluation isn't just for special projects. It should be incorporated into your organisation's everyday workload. The form it takes will vary depending on the size, nature and even culture of your organisation, but the general principles outlined here are equally applicable to the health service with an annual budget of \$5 million and the scouting group that runs on the smell of an oily sausage sizzle.

A systems approach

An organisation has a responsibility to ensure that members of staff are properly prepared for the work they do, with the appropriate skills, qualifications, licences, permissions and police checks. This applies to volunteers as well as paid staff. But you shouldn't need to reinvent the wheel every time you want to recruit a new employee or advertise for another intake of volunteers. A robust recruitment framework sets out systems and procedures for advertising, preparing selection criteria, conducting interviews, evaluating candidates, checking references, and inductions.

Similarly, evaluating the performance of paid staff and (where appropriate) volunteers should happen as part of a system in which every individual understands their role, their responsibilities, and the criteria on which their performance in that role is appraised.

The same applies to evaluating and managing risk within an organisation. A culture of evaluating and managing risk should permeate an

No examination of a project's success is complete without looking at the performance of the organisation that is running the project. The results of such an examination might not be offered externally, but the appraisal is an important exercise for a not-for-profit organisation to undertake nonetheless.

Aspects of organisational performance such as management and process evaluation are just as important as those that focus exclusively on the project at hand – maybe even more so.

organisation from the board (“What are the risks associated with borrowing \$200,000 to fix the drainage problems on the footy oval?”) to the roster of volunteers who clean the toilets after each match (“I’d best not leave the bleach where the under-five players might find it – common sense tells me it’d be dangerous, and besides, it’s covered by our hazardous materials policy”).

Taking a systems approach to evaluation – evaluating candidates, on-the-job performance, risk, the budget, the finances, evaluating whatever you care to evaluate, and doing it carefully, logically and thoughtfully – will help you to:

- progress your mission
- look after your volunteers and paid staff
- know what’s working and what’s not, and why
- be aware of unintended outcomes
- adapt as needed.

If after all that you’re thinking, “But I just want to get on with the work”, then be assured that you’re probably already practising evaluation. Every time you reflect on why your client was frustrated at today’s meeting, or think about how you might make sure that the annual fundraiser doesn’t go under in the rain like it did last year, or survey your volunteers to see what they like about working for your organisation, you’re evaluating. Taking a systems approach just means understanding things as part of a bigger picture.

Process evaluation

Are we doing it right? Process evaluation involves examining the way you carry out a task, with the aim of seeing whether you can do it better. This means that you must carry out process evaluation while you're still actually doing the job, so that the results can be fed directly into your procedures. For this reason, process evaluation is sometimes called "continuous evaluation" or "continuous improvement".

Process evaluation is necessarily short-term, because you want to know what you're doing wrong (or right) in time to do less (or more) of it.

Your project plan – recorded in your diary, or on a Gantt chart, or in your database – should include intermediate targets, showing where you expect to be at certain points along the way: your project milestones. Each of these points forces you to ask whether your efforts are going to be sufficient to reach your eventual goal or whether you will have to make changes – to bring in new resources or to change your methods. If you are falling behind, it is important to inform all stakeholders at this point. Funders may dislike bad news, but they dislike nasty surprises even more.

If you're going to modify your program in real time you need evaluation methods that are simple, quick, and definitive. Ideally, you will have planned to run your program in several separate sections, enabling you to find quickly which form of email, for example, has the best response rate, or which teaching format is better at client retention, or which training method is the most productive.

Could it be done more cheaply?

Cost-effectiveness is the relationship between inputs and outputs (or outcomes, depending how sophisticated your analysis is), expressed in dollar terms. The measure is cost per percentage point of improvement, and the variables are the extent to which a program's inputs can be minimised for a given level of program outputs, or outputs maximised for a given level of input.

When you begin putting your project plans into practice you may find that some of the elements you originally provided for are unnecessary or undesirable, freeing up some resources for use elsewhere. You may find that your staff are able to handle the work faster and more efficiently than expected, or that they get faster over time, enabling you to extend your reach. You may find that your suppliers can lower their prices, or that new suppliers enter the market.

You may also find that some or all of these changes will work against you, cutting into your budget bottom line rather than improving it. You may also find that your timetable slips or that some elements on which you have relied become unavailable. You must have in place systems that allow you to work with your funders and stakeholders to resolve these issues.

What do the users think?

Every project needs a procedure that allows people who are dissatisfied with your operations to let you know that they're unhappy.

More broadly, there should be, wherever possible, a mechanism for seeking the views of your clients, your users, your research subjects, or your audience, giving them a broad licence to assess your contribution to their welfare and to make suggestions about your approach.

What do the staff think?

It's important to have staff and volunteers on board with evaluation. They need to be able to see the purpose of whatever questions they're being asked (which is a good deal easier if they have been involved from the beginning) and to trust the evaluator to make good use of the data they collect. They need to have reason to trust that the data won't be simply filed and forgotten.

Feedback loop

Your systems of management and governance must be responsive enough and flexible enough for program evaluation data to be channelled to a level where changes to the program can be weighed up and decided on. Procedures for incorporating the results of process evaluation into ongoing project practices should be determined in advance.

Output evaluation

Did we keep our promises?

At the end of the project you will need to be able to say whether you have done what you set out to do. In the contract with your funder you promised to produce 100 widgets. Your plans set out how you would reach your target; you monitored the production process over the required time; and now, at the end of the process, you have

reached your goal, or exceeded it, or fallen short.

Output evaluation is one of the easier forms of evaluation. Here are some of the questions you should be asking when you evaluate your project's output.

Outputs: what have we done?

A widget factory should be able to measure the number of widgets that have come off the production line. Your organisation may produce items that are less concrete – lectures, soccer games, radio programs, therapy sessions – but you still need to be able to say with reasonable certainty how many “widgets” you have produced. These are your outputs.

Your output measures – the things you plan to count – and your estimates for each of these should be laid out in your project plan. If there are any major divergences between what was planned and what has been delivered, then you should, if possible, explain why in your evaluation.

Sustainability: will it keep going by itself?

The most desirable sort of project is the one that, once started, can continue successfully on its own when funding and support are withdrawn. Ideally, the community becomes involved, individuals are enthusiastic, other resource providers emerge, and institutions are convinced of the importance of it all.

This is, however, a big ask. Many projects reach the stage where the enterprise can continue if the funding continues, but far fewer cross the bridge to full independence.

Generalisability: can we do it again?

The very nature of a project is that it is a seemingly new and enlivening set of activities that your staff/volunteers/clients respond to as separate and distinct from boring old regular work. Thus your project's subjects may be influenced by what is called the Hawthorne Effect or the observer effect, where people improve their behaviour just because they know they are being observed.

This means you have to factor into your evaluation the possibility that your staff/volunteers/clients are on their best behaviour and may not be able to work at that level over the long term.

You must also investigate whether special circumstances existed that made things work particularly well in this particular instance but that may not recur.

Scalability: can we roll it out?

Considerations about extraordinary circumstances also come into play when you are deciding whether it will be possible to reproduce the project on a larger scale. Will it be possible to staff a larger runout with employees/volunteers who are properly trained and sufficiently enthusiastic? Is there an adequate number of trained professionals in this speciality, or would any increase have to wait until more students can go through the relevant courses and graduate? Is there a ceiling on available funding? Are the political stars aligned? Even successful small-scale performance is no guarantee of general applicability.

System improvement

Is there any scope for improvement in the operations of your widget factory? Seek out participants' feedback and assess whether any change in policies, procedures or structures might have improved your project's output.

Outcome evaluation

How has the situation changed?

Outcome evaluation describes the situation at the end of the project, compares it to the situation at the beginning of the project (the baseline) and maps the changes (or lack of changes).

Changes might be observed

- in the participants
- in the available resources
- in organisations and structures.

Your funders and other stakeholders are going to be very interested in what you have to say here, so communicating outcomes is key to maintaining good relations with them.

Changing attitudes

Your action may have changed attitudes in the participants, or the region, or society. You may have reinforced desirable attitudes or diminished undesirable ones (discouraging cigarette smoking, perhaps, or increasing awareness of the consequences of climate change).

There are a number of ways such changes may be measured against a baseline, among them surveys, examination of media coverage or continuing activity.

Building capacity

You may have increased the level of knowledge of the relevant facts in the participants, or the region, or society, and given people the means to change their lives in response. Their participation may have taught them some lessons about collective social activity to bring about change, and they may have developed technical skills, social skills, or bureaucratic skills that will assist them on similar projects in the future – and will allow them to mount projects of their own.

Because this is a process skill, it is more difficult to measure than changes in attitude. The real test is whether people behave differently in the future.

Changing structures

It is also possible that your project will leave behind mechanisms that will facilitate the next step – continuing associations with a common goal, procedures for consultation, links between the community and local, state or federal governments, precedents that will guide people working in the field.

Any changes of this nature should be recorded centrally and taken into account in the final balance.

Impact evaluation

How is the world different?

Chinese statesman Zhou Enlai was once asked what he thought the effects of the 1789 French Revolution had been. “Too early to say,” he replied.

Impact evaluation is the highest level of the evaluation hierarchy, and measures what a project has left behind it that is sustainable in the long term. It is by a long way the hardest thing to measure, and the most confusing and ambiguous.

The simplest problem with long-term evaluation, though by no means the least, is the length of its term. It is generally very difficult to find resources sufficient to track a project's effects over any extended time. Funders are unwilling to commit funds years in advance for later evaluation procedures. Boards and stakeholders change their priorities or move on

and are replaced by others.

Most organisations have time horizons that extend to the end of their current strategic plan, if that. Keeping a focus on the consequences of what was done last year or last decade requires a wholehearted commitment by everybody concerned, and is thus rare and valuable.

What's even more difficult is proving that any particular project or intervention created the change.

How much has the world changed?

To find out what effects your work has had, you must begin by seeing what, if anything, has altered permanently since you began in regard to the issues you sought to address. Now that the funding has gone away, what remains?

Is the population you have been working with happier? Or healthier? Richer? Better informed? Better behaved? Better organised? Better equipped? Better housed? More skilled? More generous? More tolerant?

What did you originally predict, or project, the situation would be at this time? How does the vision compare to the reality?

How much of the change was due to us?

Having measured change from your baseline, it is then necessary to assess how much, if any, of that change was due to your own efforts, and how much was due to other participants, other factors, or general societal change.

The problem is that nothing in this world has only one cause. A number of factors come together to produce change – contributions that may seem minor but without which the thing would never have been possible; contributions which seem large and important but could just as well have been replaced by other activities; contributions that moved in step with the project but did not actually influence it. Correlation does not equal causation. Let's say that again – correlation doesn't equal causation. Just because two things happened in step does not mean that one caused the other. If you still need convincing, have a look at the spurious correlations charted at <http://tylervigen.com/spurious-correlations>.

In nuclear physics it is impossible to be certain about a particle's position and its velocity. The more exactly you measure one, the vaguer you must be about the other. A similar problem affects evaluation. The more precise you can be about the nature of your project's long-term

effects, the less important that effect probably is. The more people your project affects, the more people affect your project. Your project can have a large effect on a comparatively small matter, or a small effect on a major issue; and the larger the issue you address, the harder it becomes to measure your own effect.

Any particular anti-smoking campaign, for example, could not simply point to a reduction in smoking rates as a sign of its effectiveness. Smoking rates have been falling steadily in Australia for decades, and the changes observed in your area may well be no more than the local effect of that general trend. To detect a reliable program effect it would be necessary to show either

- a greater reduction in the area affected by the project than that seen in the national or state figures; or
- a greater reduction in the area affected by the project than that seen in a designated control group.

It may also be necessary to compare your findings not to general averages but to the averages in particular groups. Smoking, for example, is more common in poor people than in rich people, and if you addressed your program to one group and then measured your success in the other you could seriously distort the actual situation. In certain circumstances it may be appropriate to correct for socio-economic status, ethnicity, English language skills, and age.

Providing a control group for social interventions is complicated, expensive, and thus not often done, but is almost irreplaceable if you wish to attain reliable information about impacts.

Statistical procedures exist for calculating how large a control group is needed for a predicted effect size and how much representation there would need to be for sub-groups.

If it's beyond your group to carry out full control procedures (and it will be beyond most groups, for sure), different kinds of controls may also be used; for example, half the group might be put through the program at a time, with measurements being taken from both groups at the start, at the end of the first stage, and after all have gone through.

Making room for unplanned outcomes

If it is hard to prove that your work alone has achieved your goals, it is even harder to say what changes you have brought about that you did not anticipate. There are many causes of every change, and there are also many consequences – some desirable, some less so – of every cause. Many organisations and many funders are quite happy to ignore these

unplanned alterations. We decide in advance what it is we want to know, what we regard as important, and what is going to be recorded.

Most evaluations don't deal with what actually happens during a project. They deal with a very much smaller subset: the things that happen in the areas we thought were important before we had the benefit of actually doing the project.

Some forms of evaluation are better at eliciting tangential developments than others. Canadian health promotion researcher Ron Labonte, for example, favours a methodology that is specifically interpretive, where the research findings arise in the course of the process of inquiry and rely on "iteration, analysis, critique, reiteration, reanalysis, and synthesis" – or, to put it another way, talking it through with the players. Labonte describes this as the story-dialogue method, operating through a reflective dialogue between participants.

The project evaluation process might in this format begin with both the practitioners and the community being asked to write up their stories about what happened during the intervention, stories that come from their own experience and their own understanding. The next step is a workshop where all the parties collect to discuss these case studies, breaking into small groups around each storyteller for intense reflection. One person acts as recorder, and the group feels its way to insights into the situation. The aim is to generate a deeper understanding of the underlying themes and to identify the tensions that exist, and the dialogue is structured to move the discussion from a simple description of what happened to some sort of explanation, or explanations, of how it happened.

At a more base level, some funders now are asking not-for-profit organisations to include in their final project report details of any unintended or unexpected outcomes. This is certainly better than nothing.

Counterfactuals

What would have happened if we hadn't been there?

If we wanted to establish scientifically how much, for example, The Wilderness Society has contributed to the cause of the environment in Australia, it would be necessary to run the last three decades of Australian history over and over about 50,000 times with variant hypothetical inputs and see what difference it would have made if the Society had never existed, or had behaved in other ways. Would other people have done the same work? Would the Franklin Dam have been built, or was the culture moving against ecological disruption with or

without the Wilderness Society to push things along?

If we were discussing the contribution of a particular anti-smoking project to a decline in smoking rates, we would need to acknowledge that an enormous cultural movement against risky behaviour and in favour of healthier activities might well have swamped any observable effect from our own program. On the other hand, that societal movement is at least partly the product of our project and others like it. Our work both feeds into and takes strength from changes in our society.

Clearly, we cannot rerun history, and there are no conclusive answers to such questions. There are, however, more convincing and less convincing stories. We cannot measure the development of situations in the real world, but we can and do try to understand them, and we can try to trace the threads of our actions forward and backward in time, sketching in likely links where the colours are obscured or the path is impossibly tangled.

Gathering the data

The data collection method you choose will generally be dictated by what you wish to measure, and how you wish to measure it.

Try not to over-complicate your data collection – but don't over-simplify it either. Work out what you need to know, then find the methods that will deliver that knowledge.

Ideally, you want your measures to be SMART. SMART is an acronym for five criteria:

Specific

The more specific you can make your aims, the easier it will be to measure them. This approach favours goals that can be expressed in numbers – dollars, percentages, client numbers. The objective is clearly defined.

Measurable

Essentially, this calls for information that can be quantified. If your objective genuinely cannot be reduced to quantities, this raises many questions about whether it is a valid aim. Almost all activities are measurable at some level.

Achievable

Your expectation of what can be accomplished must be realistic given the time, the resources and the obstacles you are dealing with.

Relevant

In their quest for measurability and quantification, people will sometimes settle on project measures that are hard-edged and definite but not closely related to what they actually want to achieve. This should be avoided.

Time-bound

Any objective that doesn't have a due date is effectively worthless.

Quantitative data

Evaluators tend to like to measure items that can be counted rather than estimated, and observed rather than deduced. Quantitative research gathers data in numerical form which can be put into categories, or in rank order, or measured in units of measurement. This type of data can be used to construct graphs and tables of raw data.

The easiest scale to work with is money. If issues can be expressed in monetary terms this means that they are

- depersonalised – within broad limits a dollar has the same value in the hands of any person, in any place, at any time
- linear – the tenth dollar, or the hundredth, has exactly the same value as the first
- convertible – a range of other values can be converted more or less successfully into monetary terms (for example, some health promotion agencies work on the basis of a standard value for a year of healthy life)
- ranked – there can be no argument about whether \$150 is better, greater, or more desirable than \$50.

Other measures may lack one or another of these characteristics.

Quantitative measures are generally seen as objective. Subjectivity, however, enters the picture with the choice of measure – a decision about what is important in the project.

While quantitative data may not be contestable in itself (the number of cars passing an observation point, for example, will be the same for any observer), the implications of a numerical measurement, and the recommendations to be drawn from it, do not arise automatically from the data and may involve as much speculation as qualitative measurements.

Qualitative data

Working with qualitative data involves the exercise of judgement. It requires understanding and describe differences rather than converting them into numbers. Subjectivity may be unavoidable. Empowerment, for example, is not an object that exists separately from the internal states of the people who experience it, so it can be difficult to measure.

Milestones

Baseline measurement

The first stage in any evaluation is to measure the baseline so that it is possible later to know what has changed. Without a baseline, it is very difficult to establish when any development occurred, and thus even more difficult to establish what caused things to change.

Input measurement

In the course of your project, you must record all relevant inputs – the number of hours spent on the project, the employees or volunteers responsible, the costs incurred. The process for recording these items

should be widely known and the person or persons responsible for carrying it out clearly identified. You should be able to draw on this database to analyse any relevant correlations.

Output measurement

As your project runs its course, you should also record all the measurable outputs you have delivered: the number of seminars you've run, the number of bales of hay you've given to rescued donkeys, the number of games your team has played. Again, the process for recording these items should be widely known and the people to carry it out clearly identified.

Endpoint measurement

At the close of your project you will be called upon to take stock of your completed efforts, to decide whether they are satisfactory, and to extract from your conclusions those factors that will influence your future planning.

Data collection tools

Interviews

An interview is a face-to-face discussion between someone from the project team and someone from the project population. Interviews are quick, simple, and comparatively cheap. You can ask about facts or opinions – and the answers you get will depend on how much the person knows and how much they're prepared to tell you.

Interviews can be structured (where you ask everybody the same questions), semi-structured (where you follow a set of flexible guidelines), or unstructured (where you go wherever the discussion takes you). The more you think you know about the situation, the more structured your interviews will be. The more structured your interviews are, the easier it will be to get quantitative data from them.

Interviews are only cheap if you don't do too many of them, and small numbers may mean that they are misleadingly unrepresentative unless you combine them with other data sources.

Focus groups

A focus group isn't just an oversized interview, or an instant survey. It's a way of seeing how people's views develop in a social context. The moderator has to provoke a dynamic interaction between participants. Focus groups are harder to manage than interviews, and harder to record, but they can help you to divine how a community will think after it's had a bit of time to reflect.

Because a focus group is a social process, it involves the same social issues you encounter in the field. If you have the bosses and the workers in the same group, for instance, neither is likely to be entirely frank with you in front of the other.

Surveys

A survey involves collecting standardised data from a sample of the target population. It quantifies information according to established statistical rules.

Respondents provide the data needed by filling out a questionnaire. Most questions should be tightly structured, but a few unstructured ones may be included (yielding, it is hoped, answers that are more expansive but less amenable to statistical analysis). Alternatively, you can limit the data range by including multiple-choice questions where the respondents must select one of the answers you provide. A well-structured questionnaire helps you to generate statistics from your data.

Effective surveys are neither simple nor easy. They require specific and professional design skills. They take some time to set up, and some time to run. Their value depends very largely on how good you are at getting a sample that's representative of the total population. Your sample has to be randomly selected, but must also be broadly representative of the sampled population (a general population sample that was 90% male, for example, or 50% Buddhist, would raise questions). Getting a sample that's large enough to produce significant results is often difficult. Be prepared to invest considerable resources.

If you want to measure:

Changes in numbers

Changes in awareness, knowledge, or skills

Increases in the number of people reached

Policy changes from your advocacy

Changes in behaviour

Changes in community capacity

Changes in organisational capacity

Changes in health/income/education among particular populations

Changes in quality of life

Data sources could include:

- counts
- measurements
- tests
- questionnaires
- surveys
- documentation
- surveys
- proxy measures
- surveys
- historical analysis
- case studies
- surveys
- demographic data
- surveys

Listening to the data

Whatever your chosen data collection method, problems will almost inevitably stand in the way of entirely clear answers to your questions.

You may have to make allowances for inadequate data (you may not, for example, be able to survey everybody involved), missing data (not everybody you invite to participate will respond to your survey), and misleading data (not everybody who responds will answer your questions honestly).

Nonetheless, some data is nearly always better than no data at all.

Dealing with complexities

Life wasn't meant to be easy, and neither was evaluation. But don't let the complexities turn you off from having a go.

Evaluation should guide change for the better. There are, however, many reasons why this does not always happen – and many examples of evaluation that has proved unproductive or actually harmful.

Misinterpreting the data

Numbers can give a misleading picture. Evaluation can be

- inappropriate, where the measures chosen do not properly represent the project goals;
- irrelevant, where the measures chosen are unrelated to the project goals.

Fiddling the data

Unfortunately, project evaluation is often distorted by an organisation's need for continuing funding. Difficulties or shortfalls may not be reported in case they create doubt in the mind of the funding body responsible.

Where disclosure is unavoidable, difficulties are often minimised or misattributed. Straightforward lying is probably rare, but where different measures give different results, the most favourable will be put forward, and where ambiguities exist they will invariably be resolved in favour of the incumbent.

Funding bodies themselves often encourage the concealment of mistakes by defunding applicants who have acknowledged past failures, rather than by working with them to incorporate the lessons learned in the next iteration of the project.

Teaching to the test

If the wrong measure is chosen, managers can work to the measure rather than the goal. Under the old Soviet system, for example, factory managers were rewarded for increasing production, not for increasing sales, which meant that canny managers would churn out 10 years' supply of unsaleable items and simply leave them on pallets in the shed. If teachers are assessed on how many of their students pass a particular test, they may respond by training them in the specifics of that test while neglecting other important educational areas – “teaching to the test,” as they say.

The evaluation can distort the program.

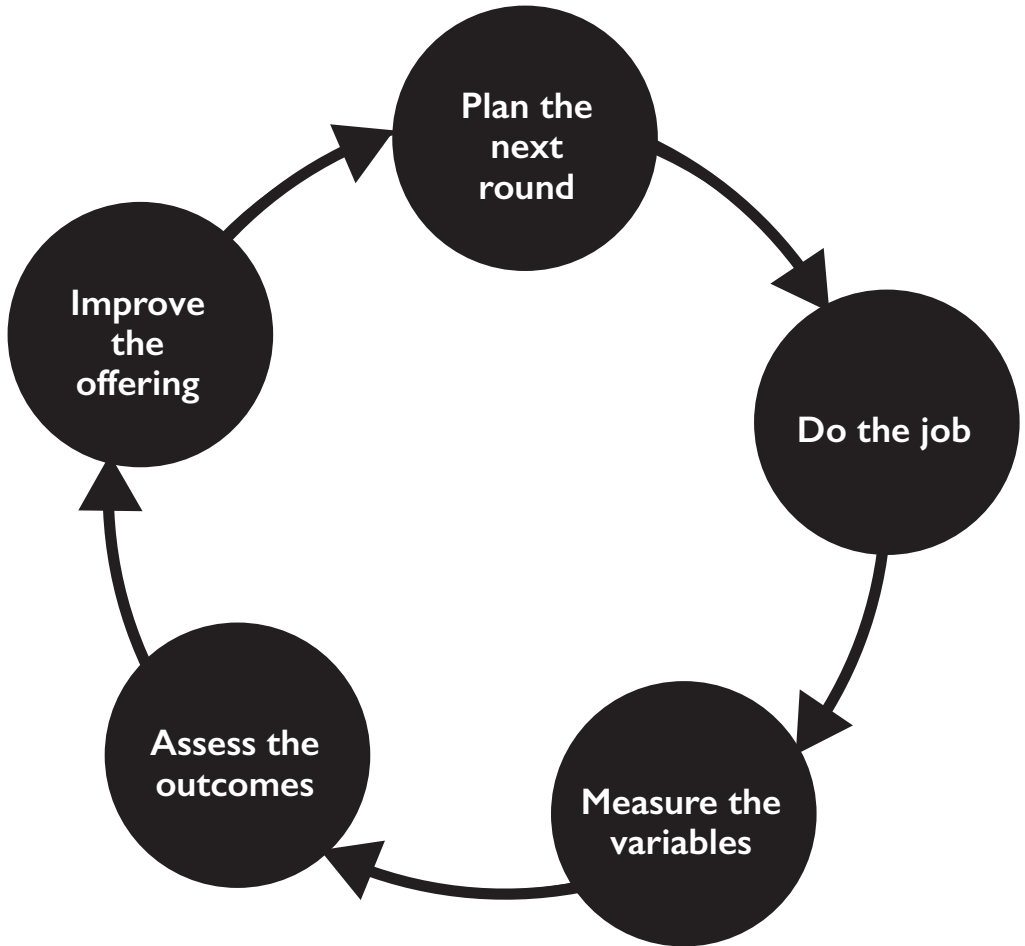
Wasting resources

Sledgehammer, meet nut. Evaluation can be

- disproportionate, where small, cheap projects are burdened with extensive reporting requirements that consume resources essential for the program. While no project should be allowed to end without any evaluation at all, that evaluation should be appropriate to the project scale, and might at one end of the spectrum consist of no more than half a page of written impressions.
- unused, where measurements called for simply by habit are taken and filed away without being analysed. Any evaluation that will not be incorporated into the decision-making process should not have been carried out in the first place.

The evaluation cycle

Evaluation is not a one-off activity (or, at least, it shouldn't be). Evaluation should be part of a continual cycle of learning and adaptation. The cycle looks something like this:



Every successive iteration should do better.

The term “better” is, of course, subject to interpretation. It can mean cheaper, or larger, or prettier – improved financially, or quantitatively, or qualitatively.

And, because anything that changes the world has to interact with everything else in the world, there are issues of time and chance.

You'll have to find resources to fund the evaluation. Then you'll have to decide whether the evaluation was worth what you paid for it – you'll have to evaluate the evaluation. And then you can start all over again.

Sharing your lessons

Understanding more about the outcomes and impacts of your work is invaluable. But don't stop there.

Consider sharing your results with grantmakers and others who fund similar work, as well as other organisations – even “competitors” – that could benefit from that knowledge.

Try not to gloss over the “bad” stuff – some of the best lessons are derived from hearing about what didn't work.

Include unintended outcomes too, where possible.

By sharing your evaluation results – the good, the bad and the ugly – you'll be magnifying the results of your project.

EIGHT WAYS TO ENHANCE EVALUATION IMPACT

Evaluations That Make a Difference is a collection of eight stories about project evaluations that made a difference not only from the perspective of the evaluators but also from the perspective of the commissioners and the people who were the subjects of the projects. You can read more about the stories on page 6 and online at <https://evaluationstories.wordpress.com>.

In addition to drawing out the eight stories, this project teased out the “enabling factors” that contributed to the impact of the projects behind the stories.



The project found that evaluations that contribute to social betterment share the following characteristics. They:

- Give voice to the voiceless – all voices need to be heard, but those without power often go unheard
- Provide credible evidence based on excellent design and methodologies
- Use a positive approach emphasising strengths, opportunities, aspirations and results
- Actively engage users and intended beneficiaries through a utilisation-focused process that gets buy-in as the evaluation progresses
- Embed evaluation within the program right from the start, if possible, in order to have base-line information and to promote evaluative thinking
- Sincerely care about the evaluation so that commissioners, users and evaluators are working together to ensure credible evidence
- Champion the evaluation – evaluators need to work with commissioners and users to help them understand how the evaluation can contribute to making decisions
- Focus on evaluation impact – from the beginning evaluators need to think about the potential effect of the evaluation on the program and program participants.

Sample evaluation reports

What form should your evaluation report take? Should it be a single page on your website, a 120-page commercially printed full-colour report distributed to all your funders, staff, volunteers, members and project participants, or something in between?

There's no right or wrong answer to this question, but you probably won't be able to cover all the ground you need to cover in a single page on your website, and there's nothing intrinsically better about a lengthy report.

Our advice is to consider your intended readership, consider the most compelling, succinct way to convey the information you need to convey, and write your report in plain English – no jargon, no weasel words, and no more acronyms or abbreviations than absolutely necessary.

If you've planned and executed your evaluation thoughtfully, and if you stick to plain English when it comes to writing the report, you'll be on a winner.

Below we've listed some examples of published evaluation reports. Not all of them meet all those criteria (there's not much value, for example, in publishing a pie chart for each one of your data sets without analysing the data and explaining why it's significant), but examining flawed evaluations can be as instructive as looking at truly insightful ones.

Family by Family, Australian Centre for Social Innovation

<http://www.tacsi.org.au/wp-content/uploads/2014/08/TACSI-FbyF-Evaluation-Report-2012.pdf>

Keep Watch @ Public Schools program, NSW

http://www.royallifesaving.com.au/___data/assets/pdf_file/0004/10966/RLSNSW_KWPP_Phase2_Report.pdf

Macarthur Youth Mental Health and Housing Project

http://www.neaminational.org.au/sites/default/files/sprc_report_myp_evaluation_final.pdf

Tooty Fruity Vegie Project

http://epubs.scu.edu.au/cgi/viewcontent.cgi?article=1139&context=educ_pubs

KidsMatter Early Childhood Evaluation Report

<https://www.beyondblue.org.au/docs/default-source/research-project-files/bw0077-kmec-evaluation-full-report.pdf?sfvrsn=2>

National Anti-Racism Strategy and “Racism. It Stops with Me”

https://www.humanrights.gov.au/sites/default/files/document/publication/WEB_NARPS_evaluation_2015FINAL.pdf

Skills for Education and Employment Programme Evaluation

https://docs.education.gov.au/system/files/doc/other/see_programme_evaluation_report.pdf

Evaluation of the Northern Region Prevention of Violence Against Women Strategy

<http://www.whin.org.au/images/PDFs/PVAW%20Evaluation%20REPORT%202013.pdf>

Aboriginal Family Violence Prevention and Legal Service Victoria’s Early Intervention and Prevention Program

<http://www.fvpls.org/images/files/Evaluation%20report%20EIPP%20Document%20REV%20WEB.pdf>

Connected Communities Strategy – Interim Evaluation Report

https://www.cese.nsw.gov.au/images/stories/PDF/Connected_Communities_Interim_Report.pdf

Further reading

There is a sea of information about project evaluation available online. Here are our picks.

Website: Better Evaluation

<http://betterevaluation.org/>

An excellent website by an international coalition of evaluation professionals and academics, including Australians. The section headed "Approaches" highlights the fact that for every evaluation challenge, there is an approach to suit. For example, if you're finding it difficult to prove causality when evaluating your program, try a contribution analysis.

PDF booklet: Measuring Outcomes

<http://strengtheningnonprofits.org/resources/guidebooks/MeasuringOutcomes.pdf>

A 55-page PDF booklet – a little dry, but well structured and comprehensive. Written for the US market but easily transferable to the Australian context.

Web page: Basic guide to program evaluation (including outcomes evaluation)

<http://managementhelp.org/evaluation/program-evaluation-guide.htm>

A low-tech but highly readable guide from the US-based Free Management Library.

More titles from Our Community

Choose from a great range of practical books written by Our Community to help you and your organisation. Learn more and place your order at www.ourcommunity.com.au/books.

.....

Find grants and raise money

Fire up Your Fundraising Events: How to Make More Money While Having Fun (includes online templates)

The Complete Community Fundraising Handbook: How to Make the Most Money Ever for Your Community Organisation

The Complete Schools Fundraising Handbook: How to Make the Most Money Ever for Your School, Pre-school or Kindergarten

Winning Grants Funding in Australia: The Step by Step Guide

How to Find Money Fast: 50 Great Ideas to Raise \$5,000

Simple Secrets of Successful Community Groups Volume 1: Over 400 Tips on Running a Successful Community Group

Simple Secrets of Successful Community Groups Volume 2: Another 400 Tips on Running a Successful Community Group

How to Manage Your Grant after Winning It! (includes templates on bonus CD)

Great Fetes: Fundraising and Fun – Without the Fuss (includes templates on bonus CD)

More than Money: How to Get Money, Resources and Skills from Australian Businesses

.....

Build a better board or committee

The Board Doctor: Expert Diagnosis for Board & Committee Ills

Making Meetings Work: Conquering the Challenges and Getting Great Results

Surviving and Thriving as a Safe, Effective Board Member: Facts You Need to Know Before, During and After Joining a Community Board

Transforming Community Committees & Boards: From Hell to Heaven

Get on a Board (Even Better – Become the Chair): Advancing Diversity & Women in Australia

.....

Promote your organisation

How to Stand Out from the Crowd: The Complete Marketing & Media Handbook for Community Organisations

Effective Letters: 50 of the Best Model Letters to Help Community Organisations Fundraise, Connect, Lobby, Organise & Influence