# Assessing Science Communication Effectiveness: Issues in Evaluation and Measurement

**Rod Lamberts and Catherine Rayner**

**Abstract** The practice and theory of evaluation is far too large a topic to comprehensively cover in a single book chapter. We present here a subset of that topic which considers evaluation in a science communication context. Between us, we draw on some 35 years' experience in social research and evaluation which has given us insights into what works and what often gets missed. We will touch on some fundamental elements of evaluation but focus more on offering critiques and pointers that we have picked up through our professional endeavours. We will also touch on the political nature of evaluation, particularly in science communication and in the evaluation of university research and education.

With this in mind, this chapter moves from the general to the specific. We begin by presenting and critiquing some fundamental concepts and tools of evaluation before turning to specific challenges facing the evaluation of science and communication endeavours. Examples include a classic evaluation controversy that spawned the Public Understanding of Science (PUS) movement in the UK and a look at the difficulties faced by science centres in determining the effect they may be having on their visitors. As an example of a sector-wide evaluation challenge, we look at an issue of broad concern to science communication because it affects science research in general: the global ranking of universities.

At the end of the chapter, we have included a more detailed, but not exhaustive, example of a hypothetical evaluation of a science communication enterprise.

R. Lamberts (✉) • C. Rayner
The Australian National Centre for the Public Awareness of Science, The Australian National University, Canberra, ACT, Australia
e-mail: Rod.lamberts@anu.edu.au; Catherine.rayner@anu.edu.au

L. Tan Wee Hin and R. Subramaniam (eds.), *Communicating Science to the Public*,
DOI 10.1007/978-94-017-9097-0_8. 

## What Is Evaluation?

At its most basic, evaluation is about establishing the value or worth of something. More precisely, it involves systematically and rigorously gathering data that support decisions about the effectiveness of products or activities and that offer a base for improving those products or activities (these will usually be referred to as 'enterprises' throughout this chapter).

Broadly speaking, there are two main types of evaluation: summative and formative (see Fig. 1). Summative evaluation (also called 'ex-post evaluation') is more common and tends to focus on the impact of an enterprise after it has finished. Formative evaluation refers to evaluation efforts taking place before or during the implementation of an enterprise with the aim of assessing performance, and perhaps improving it, along the way.

Evaluation begins with the setting of clear and explicit goals, objectives and priorities for the enterprise under scrutiny. Without these it is impossible to move ahead as there is nothing against which to measure success. Put simply: you can't determine what you've done if you don't define what you set out to do.

Whenever people try to set goals, objectives and priorities, there will necessarily be negotiation, prioritisation and compromise. Different groups involved in an enterprise may well have different goals and, with these, different criteria for measuring success. For example, those providing funding may have quite different priorities to those working at the coal face, which may be different again to the priorities of the people for whom an enterprise is intended. At its core, evaluation has inherently political dimensions.

As multifaceted as evaluation can be, it doesn't have to be complex to be useful, but to be effective – and credible – it must be scrupulously designed, constructed and executed. And all this must be done with awareness of the greater contexts in which the enterprise to be evaluated operates and the context in which the evaluation itself will be conducted, analysed and reported.

## *Is Evaluation Research?*

Quite often, particularly in the higher education sector, the question arises as to whether what you are doing is 'evaluation' or 'research'. It is easy to argue that many evaluations meet the OECD definition of applied research:

> Applied research is original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective. (OECD, 2002, para. 64, p. 30.)

The distinction ultimately depends, once more, on context. Each enterprise being evaluated, each organisation requesting that an enterprise be evaluated and each

team of investigators conducting an evaluation will have an opinion about where research starts and ends.

## Why Evaluate?

There are really only two reasons to evaluate. For the purists, the first – and arguably *only* – reason is to assess the extent to which an enterprise met its goals and objectives. Such assessment can then be used to modify successive implementations of that enterprise with a view to continuation, improvement or cessation.

The second reason is that there may be an obligation to undertake something that appears at least 'evaluation-like' for reporting purposes (often referred to as 'box-ticking' exercises).

These two reasons are by no means mutually exclusive, although evaluation driven merely by a requirement to report can run the risk of being less than informative.

Regardless of the underlying reasons, effective evaluation can help people:

1. *Understand* what an enterprise achieved using 'objective' measures and methods
2. Find ways to *improve* the enterprise
3. Usefully *communicate* the extent to which an enterprise met its goals and objectives using methods and measures that are broadly intelligible to people beyond those involved in the enterprise itself
4. Provide evidence-based information to *support* ongoing – or persuade new – funders, backers or participants

Skilfully and mindfully conducted, the process of evaluation can itself improve people's impressions of the enterprise under consideration. In addition, the results of a well-designed, conducted, analysed and communicated evaluation can be enormously persuasive. This is particularly useful when seeking to maintain, or generate new, sources of support.

## What and Who to Evaluate

There are innumerable initiatives that make up the suite of potential science communication enterprises, and in theory all of these have goals and objectives that could be evaluated. A science communication enterprise may include any or all of the following, all of which can be subject to formal, systematised scrutiny:

- Events
- Programmes
- Presentations
- Exhibitions
- Classes
- Training activities
- Research activities
- Performances
- Communication artefacts (such as brochures or websites)

In addition, evaluation efforts themselves can be assessed.

Of course, some enterprises, or aspects of them, are more difficult to assess than others.

In any science communication enterprise, there is likely to be a number of stakeholders, all of which can be evaluated as part of that enterprise. The kinds of stakeholders that might be evaluated include, but are not limited to:

- Participants or audiences
- Presenters, designers and creators of the material making up the enterprise
- Donors/funders
- Managers
- Public officials

In addition, the evaluators themselves can also be evaluated.

## When to Evaluate

'When' can refer to two things here. First – time, second – 'under what circumstances?' (the more political version of 'when').

Typically the latter occurs any time you want to gain, or maintain funding and/or popular, leadership or administrative support. Political considerations might also include how closely an evaluation timetable should be aligned with the political cycles relevant to the enterprise under scrutiny.

Turning to timing more broadly, many practical issues influence the timeline of evaluation efforts. These include, but are not limited to, such questions as:

- What baseline or comparison data is required?
- Whether to include formative evaluation and, if so, will it be used to modify the enterprise along the way?
- Will the timing of evaluation efforts influence the validity of the evaluation or the likelihood of getting responses?

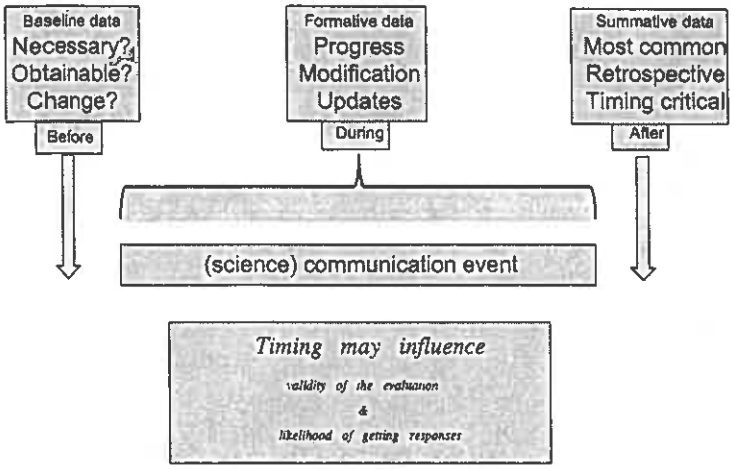


Fig. 1 Timing your evaluation

## How to Evaluate

### *Goals and Objectives*

A robust, practical evaluation cannot begin without at least having:

1. Something to evaluate (in this case, some kind of science communication enterprise)
2. Explicit, clearly articulated and measurable goals and objectives for that enterprise

These two elements underpin the key starting principle behind all effective evaluation.

In addition, goals and objectives must be appropriate for the context in which, and for whom, the enterprise and the evaluation are being conducted. An effective evaluation will be influenced and guided by:

- The goals and objectives of the enterprise being evaluated
- The people *for* whom it is being evaluated
- The people *by* whom it is being evaluated

### *Measures and Indicators*

To assess the extent to which the goals and objectives of an enterprise have been met requires measures that validly and reliably represent phenomena that explicitly

reflect the effects of these goals and objectives. Developing such measures can be difficult, and it is usually far more difficult to measure *outcomes* than outputs and activities. Because of this, we often see evaluation efforts employing inappropriate measures or, more commonly, using appropriate measures in the wrong way.

For example, counting the number of participants attending an annual science festival is straightforward. Such data offer simple *output* and *activity* indicators for the festival itself: an increase in numbers is a straightforward indicator that the event is becoming more popular. What such data don't tell us is anything about is *why* it has become more popular.

There is nothing wrong with measuring such straightforward outputs and activities as long as they are not used as surrogate measures of outcomes.

It is possible to have measures that are *reliable* – that is, they consistently measure the same phenomenon across various contexts and times – but that are not *valid*, that is, they do not accurately measure the phenomenon of interest.

Returning to the science festival example, counting people is both a reliable and valid way of determining changes in numbers of people coming to the event year after year.

Some may be tempted to use this as an indicator of something else and so might interpret an increase in attendees as suggesting an increase in interest in science. Without other information, there is no way of knowing that increases in attendance are not linked to other phenomena. This is a case of using a reliable indicator invalidly.

Possible other reasons for attendance increasing could be the following: parking has improved over the years, competing events in previous years no longer exist, or the event moved to a more populated location.

We can *reliably* count increases in numbers of people, but without other information, we cannot *validly* make claims about what increases in that count mean or why they occurred (Lamberts & Grant, 2012).

## *Tools and Techniques*

There are many evaluation tools and techniques available and a plethora of ways to administer them. Many authors have written robust and detailed texts about these. One we recommend is by Floyd Fowler Jr (the 4th edition was revised in 2009) on survey research methods from the Sage Applied Social Research Methods series (try also Foddy, 1999).

Fowler presents a wealth of information and techniques in his monograph. Of particular relevance here is the chapter on methods of data collection. In it, Fowler covers a comprehensive array of issues to consider when choosing survey strategies and follows this up with a detailed summary of the relative merits of these strategies.

The major issues Fowler (2009, pp. 70–80) considers are:

- Sampling
- The type of population
- Survey and question format
- Question content
- Response rates
- Costs
- Type of facilitates available
- The time you have for data collection
- Matters surrounding the use of computers to assist data collection

## *Resources*

As we have discussed, the principal aim of any evaluation will be to assess the extent to which the goals and objectives of the enterprise under scrutiny were met. Each of the many tools and techniques to do this has its own resource requirements. While many factors are involved in the choice of methods, in every case, resource availability will be fundamental to such choices. Resources are pivotal in setting the scope and parameters of your evaluation activity.

### People

Evaluation efforts must be supported by people who either have pertinent skills and experience to design, conduct, analyse and report on your work or have the means to learn or access such skills and experience. Depending on the methods, you may also need people who can undertake specific, and perhaps also less-skilled, tasks. These may involve responsibilities such as contacting potential interviewees or doing data entry once results start to come in. It is also important to consider the time and effort you want your evaluation subjects to contribute to the process.

### Materials and Infrastructure

Also critical to an effective evaluation effort is having the right type and sufficient amounts of physical resources. This is not merely about money (though money certainly helps). You may need particular software or a suitable venue to conduct interviews or focus groups. If there are confidentiality issues, you will need to take steps to keep data secure.

### Access to Respondents

If your evaluation requires gathering feedback on an event, activity or communication from people, then you need access to willing and relevant respondents. If there are limits on your ability to access such people, this will influence the scope, parameters and possibly also the effectiveness, of evaluation efforts

### Time

Time is perhaps the most influential resource of all. Although it is possible to counter time shortages by increasing other resources – such as money and people – in the end, time will exert an all-encompassing influence on the scope of any evaluation effort.

## Evaluation Design: Putting It All Together

Once you have a specific evaluation to undertake, you need to decide upon and document your evaluation design and implementation plan. If the work has been commissioned by a third party, they will play a pivotal role in this process.

The Centers for Disease Control and Prevention in the USA publishes a guide to writing SMART objectives (Fig. 2). This is an excellent framework for shaping evaluation activities.

The goals, objectives and priorities of the enterprise to be evaluated form the starting point for designing the evaluation. From them, you need to formulate a series of questions to address.

The questions lead to the choice of tools and techniques, with that choice further modified by resource availability. Some of the questions which guide tool choice are:

- What analyses do you want to conduct and what data will you need to conduct them?

  If you desire large-scale quantitative data about activities and beliefs, you will need a tool such as an electronic survey with predominately closed questions. On the other hand, if you want deep information about underlying motivations for behaviour, you will need to collect detailed qualitative data from a comparatively small group.
- How do you want your data to come to you?

  Do you want, for example, paper surveys, files from electronic surveys, recordings of interviews or transcripts of focus groups?
- What's popular at the moment among significant audiences/stakeholders?

  Are they more used to/interested in/persuaded by large-scale surveys, in-depth focus groups or something else? This is particularly important in the case of a commissioned evaluation.

Fig. 2 SMART objectives

| | |
|---|---|
| **Specific** | |
| Activity | Clear objectives for your activity |
| Evaluation | More specific = more measureable |
| **Measurable** | |
| Activity | Impossible to know if objectives are met unless they can be measured |
| Evaluation | Measures must be valid |
| **Achievable** | |
| Activity | Objectives should be achievable with available resources |
| Evaluation | So should the evaluation itself |
| **Realistic** | |
| Activity | Objectives clearly related to goals and realistic in the context of activity |
| Evaluation | Measures directly relate to objectives |
| **Time-bound** | |
| Activity | Objectives include time frame for achievement. A mustfor planning. |
| Evaluation | Relate evaluation to activity time frames |

Before you finalise your tools, ask yourself:

*Will the data directly and clearly provide evidence to help me answer my evaluation questions?*

A couple of good tests of this are to run a small pilot study of your chosen tools/instruments, or build a plausible dummy dataset, and run simulated analyses.

## What Happens After the Evaluation Is Over?

Evaluation does not end with data analysis and synthesis. The next, and critical, step is communicating results.

Communicating the results of an evaluation is like any other science communication activity with the same fundamental steps:

1. Identify the key audience(s).
2. Analyse the key audience(s).
   What do they want to know? How much do they already know? What do they expect you to have found? How do they prefer to access information: written material, verbal presentation, etc.? How much time do they have to spend on understanding the report?
3. Report with the critical characteristics of the relevant audience(s) in mind.
4. Follow-up.
   Find out if the report met the needs of the audience (i.e. evaluate the report).

Running and reporting on a single iteration of an evaluation is not always the end of the story. Some evaluations are the beginning of a continuing series of monitoring and evaluation activity, some raise further questions requiring follow-up.

## Examples of Evaluation

The following three examples look at different evaluations from the field of science communication. The first illustrates the importance of knowing exactly what you are measuring, the second shows the problems of finding measurable outputs and activities that relate to your goals, and the third looks at the dangers of the feedback loop between measuring what you value and valuing what you can measure.

Each example is complex and has a detailed history, all of which could – indeed have – inspire innumerable debates in the scholarly and grey literatures. Here we look at each of them as an example of some of the points we have raised about issues in assessing science communication practice.

### *Public Understanding of Science: The Early Days*

In 1989, Durant, Evans and Thomas reported on the results of two large-scale surveys of the English and American public which sought to measure and compare public understanding of science (PUS). This became a seminal moment in the history of science communication and has become a classic example of how measurement and evaluation can be far more complex and nuanced than they first appear.

In the surveys, stratified samples of just over 2,000 people representing 'the public' were asked a suite of questions under four themes:

- Self-reports of levels of interest in science
- Self-reports of levels of informedness in science

- Understanding of the processes of science
- Measures of scientific knowledge

In this discussion of evaluation, however, a critique of the fourth one, *measures of science knowledge*, is particularly informative.

The knowledge questions Durant, Evans and Thomas (1989) asked included apparently basic science questions such as whether it is true that the Earth revolves around the sun once a year. It is true, although 66 % of UK respondents and 54 % of US respondents responded otherwise. The results of this, and other knowledge-based questions, were far from encouraging or as the authors put it:

> The results make sobering reading for politicians and civil servants with an interest in science policy, for scientists with an interest in their public constituency, and for educationalists with an interest in the dissemination of scientific learning. (p. 11)

Simply put, British and American publics were measured – *evaluated* – on their knowledge of scientific facts, and they were found wanting. This finding led to enormous amounts of money being spent on raising PUS, with the UK spending billions of pounds in pursuit of this goal.

But what were these researchers measuring, and just how valid, reliable, useful or fair were their measures?

Stocklmayer and Bryant (2011) asked more than 500 Australian-based scientists from a range of disciplines to answer the same set of science knowledge questions used by Durant et al. (*ibid*). What they found was surprising.

For the true/false question the *centre of the Earth is very hot*, the expected answer is 'true'. Of the scientist respondents, 5 % said 'false' or 'don't know', compared to 12 % of the general public. One of the main reasons respondents in the scientists group gave for their 'false' or 'don't know' answers were to do with frames of reference. How hot is 'very'? Hot in comparison to what? An apparently unambiguous question turned out to have unexpected nuances.

Another question that turned out to be more ambiguous than it first appeared was: 'Sunlight can cause skin cancer' (true/false). The expected answer was again 'true'. However, 12 % of scientists answered 'false' or 'don't know' (compared with 6 % of the general public). If you were to stop your critique here, you might conclude that the general public know more than scientists. It was actually something else. Once more, the problem was in the ambiguity the question presented to science-knowledgeable respondents. Generalising across that sample, the main complaint was that more information was needed. For example, did the question only refer to the visible spectrum, in which case 'false' is the correct answer.

So even with positive intentions, excellent skills, a substantial budget and an enormous depth of experience, evaluation can go wrong. In this case, measures of scientific literacy, though reliable, were not necessarily being applied validly.

## *Evaluating Science Centres*

When the first modern science centre, The Exploratorium, was set up in 1968, its founder Frank Oppenheimer said, 'there is an increasing need to develop public understanding of science and technology' (Oppenheimer, 1968, p. 206). He believed that the purpose of a science centre would be to satisfy the need for an environment in which people could become familiar with the details of science and technology. The current vision statement for Questacon, Australia's national science and technology centre, is:

> A better future for all Australians through engagement with science, technology and innovation. (Questacon, 2012)

Although separated by almost 45 years, the goals are similar, wide ranging, complex and difficult to measure.

There is a huge body of research into learning science in informal contexts and visitor experiences in museums, galleries and science centres. Much of this research is at the level of single programmes or exhibitions or focuses on a single outcome, such as inspiring an interest in physics. In 2009 the National Research Council of the National Academies in the USA published *Learning Science in Informal Environments: People, Places, and Pursuits* (National Research Council, 2009), a detailed look at science centres which includes comments and recommendations on their evaluation and how it can be improved.

Designing a logical chain of investigation that leads from the overall institutional goal to an evaluation of a single programme or exhibition is relatively straightforward. However, synthesising the evaluations of multiple aspects of an enterprise as complex as a science centre to determine the extent to which the enterprise as a whole is meeting its high-level goals is far more complex. This problem is considered by John Falk of the Institute for Learning Innovation is his 2004 paper, *The Director's Cut: Toward an Improved Understanding of Learning from Museums* (Falk, 2004).

The possibly unsolvable question of the long-term effectiveness in instigating positive changes in attitudes to science as caused by science centres neatly elucidates how difficult evaluation can be.

It is easy enough to measure short-term attitude changes as influenced by a particular exhibit. The results of brief pre- and post-visit questionnaires for the exhibit can be compared, and assertions about short-term effects of the exhibit made.

Determining whether there are long-term enduring changes inspired by that same exhibit is far more problematic. It is virtually impossible to account for all the possible confounding factors that influenced attitude changes between the time a person leaves an exhibition and a follow-up evaluation days, weeks or months later. A recent paper from Turkey is a rare example of post-visit evaluation being done well after the visit (Sentürk & Özdemir, 2012).

Depending on what you are looking for, science centres can be very easy or extremely difficult to evaluate.

## *Evaluating the Performance of Universities*

At the time we wrote this chapter, the global evaluation and comparison of university performance had been a topic of international news for a number of years. The two annual reports that have universities around the globe holding their breath are the Times Higher Education World University Rankings Table and the Shanghai Jiao Tong Academic Ranking of World Universities.

These two tables, and how they are used, frequently have researchers, commentators and users of evaluation results locked in fierce debate over their utility and worth.

The tables use a range of indicators of a university's quality such as:

- Research papers (number published in a given period, citations, journals published in)
- Researchers (number of Nobel laureates, members of the learned societies, proportion of those holding advanced degrees)
- Research funding
- Students (undergraduate, postgraduate and research, numbers, evaluations)
- Evaluations by other universities

Much like the measures that helped inspire the PUS movement in the UK, these indicators of relative performance should at best be approached with caution, at worst, seen as so arbitrary as to be useless.

There are many critiques levelled at the use of these measures, but perhaps one of the most interesting and alarming in a consideration of evaluation is the extent to which

> the evaluation process exerts a profound influence on many aspects of the scientific enterprise, including the training of new scientists, the way in which grant resources are distributed, the manner in which new knowledge is published, and the culture of science itself. (Vale) (2012, Abstract)

In essence the act of evaluation can change that which is being evaluated. Focussing resources on trying to do well on specific measures is not always the same as doing well in the pursuits purported to be captured by that measure:

> The idea to reward those who are productive seems fine at face value, but that idea has become ideology. Metrics of quantity once were the means to assess the performance of researchers, but now they have become an end in their own right. Ironically, once individuals actively pursue certain indicators of performance, those indicators are no longer useful as independent yardsticks of what they were once meant to measure. (Fischer, Ritchie & Hanspach, 2012, p. 473)

A classic example of this is in scholarly publishing and its link to the promotion of academics in universities. The more individuals and their universities are focussed on measuring the number of scholarly papers that are produced, the more there is a danger of quantity becoming a substitute for quality. The pursuit of the easily measureable activities and outputs again drives the behaviour of the people being judged by that measure (and so changes outcomes).

There is rigorous discussion and debate about the usefulness, validity and effects of this type of indicator in a range of academic and scholarly online outlets (Aitkin 2012; Curry 2012; Donald 2012; Rossi 1999; Sawyer 2012).

## Evaluation Example: A Public Lecture on Genetically Modified Food Crops

Imagine you are conducting a public lecture for a general adult audience extolling the virtues of genetically modified (GM) food crops, with a goal of changing negative attitudes towards the use and consumption of GM food crops among this audience.

How do you evaluate this? A straightforward approach would be to ask the audience to fill out a short pre-event questionnaire that focuses on their attitudes to GM food crops, their level of comfort with eating them and the extent to which they support farmers growing them. You then collect those questionnaires, run the public lecture and ask them the same questions again straight afterwards. A mechanism for doing this would be to hand out and collect the pre-event questionnaires before the event starts and then place the post-event questionnaires in an envelope with instructions asking audience members not to open them or fill them out until after the event finished (the end of the public lecture). The completed post-event questionnaires would be collected as the audience leaves.

Although demographic information such as age of respondents and previous experience with GM could be relevant here, it is not included for two reasons: it is usually better to ask as few questions as possible to encourage more people to respond, and we wanted to keep the example short and focussed.

The extent to which people's attitudes changed between the pre-event and post-event questionnaires could be used as a measure of the effectiveness of the event in instilling positive attitude changes towards GM food crops.

Figures 3 and 4 show example sets of suitable pre- and post-event questions. The two are the same except for an additional question relating to the event itself at the end of the post-event questions.

### *Suggested Analyses*

The analyses suggested here are by no means exhaustive and are presented to give a flavour of what might be useful. We focus on looking at changes across the whole sample, rather than individual pre and post differences.

Readers experienced with statistical analyses would be aware that there are a wealth of analysis options available. For readers with little, or no, such experience, we suggest seeking advice from experienced people or referring to the following

Qu 1: In general, how safe do you believe it is to eat foods that contain genetically modified ingredients? Please circle one option

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| not at all safe | | sometimes safe, sometimes unsafe | | completely safe |

Qu 2: Have you ever eaten foods that contained genetically modified ingredients? Please tick one option

- ☐ Yes
- ☐ No
- ☐ Don't Know

Qu 3: Would you eat foods that contain genetically modified ingredients if the same food was available *without*genetically modified ingredients? Please tick one option

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Don't care

Qu 4: To what extent would you support famers growing GM crops under the following conditions (please circle one option for each of the questions):

a) Significantly more could be grown in the same space as non-GM versions

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

b) The GM crops did not need pesticides

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

c) The GM crops contained additional nutrients

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

d) Significantly less water was needed for GM versions

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

**Fig. 3** Pre-event questions

references in the first instance: *Statistics for Dummies* (Rumsey, 2003), *Basic Statistics for the Social and Behavioral Sciences* (Diekhoff, 1996) or *Statistics Demystified* (Gibilisco, 2004).

Gleaning a quick and simple measure of the effectiveness of the lecture could be as simple as looking at the differences in the frequencies between the pre-event and

Qu 1: In general, how safe do you believe it is to eat foods that contain genetically modified ingredients? Please circle one option

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| not at all safe | | sometimes safe, sometimes unsafe | | completely safe |

Qu 2: Have you ever eaten foods that contained genetically modified ingredients? Please tick one option

- ☐ Yes
- ☐ No
- ☐ Don't Know

Qu 3: Would you eat foods that contain genetically modified ingredients ifthe same food was available withoutgenetically modified ingredients? Please tick one option

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Don't care

Qu 4: To what extent would you support famers growing GM crops under the following conditions (please circle one option for each ofthe questions):

a) Significantly more could be grown in the same space as non-GM versions

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

b) The GM crops did not need pesticides

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

c) The GM crops contained additional nutrients

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

d) Significantly less water was needed for GM versions

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| I would not support it | | I might support it | | I definitely would support it |

Qu 5: To what extent did the lecture you just saw change your views on genetically modified food? Please tick one option:

- ☐ Overall, I feel *more positive* about GM food than I did before the lecture.
- ☐ Overall, I feel *the same* way about GM food as I did before the lecture.
- ☐ Overall, I feel *more negative*about GM food than I did before the

**Fig. 4** Post-event questions

the post-event questionnaire on the three response categories to question 5: *To what extent did the lecture you just saw change your views on GM food crops?*

The larger the number of people indicating that they felt *more positive* after the event, the more you could argue the lecture had a positive effect on audience attitudes to GM crops. Of course this is simple indicator of the effect of the lecture,

and it does not take into account how positive, negative or ambivalent respondents were before they saw the lecture.

Comparisons here between pre- and post-event questionnaires constitute a repeated measures – a within-subjects – design. That is, the same sample is being surveyed twice, rather than comparisons being made between two groups of different people (between-subjects design). The main advantage of this kind of design is an increase in statistical power, meaning smaller changes between pre and post scores are more likely to be statistically significant than for a between-subjects designs. This also means that statistical analyses must be suitable for such a design (for more on this, see Diekhoff, 1996 or Tabachnick & Fidell, 1996).

You could compare the mean scores on question one pre versus post using a repeated measures *t*-test (see Diekhoff, 1996). A statistically significant increase might indicate the lecture positively influenced perceptions of GM safety for consumption. Similarly, you could do this comparison between pre and post responses with one, some, or all of the mean scores for question 4a, 4b, 4c and 4d.

Pre-event responses to question two could be used to determine if differences in attitudes to GM foods, and possibly responses to the lecture, are associated with personal experience (in this case, consumption of GM food). Should a comparison of pre-and-post responses to this question reveal differences – for example, if any respondents ticked 'don't know' in the pre-event questionnaire and then 'yes' in the post-event – it might suggest these respondents learned something new about GM foods from the lecture.

Similarly, changes in pre-and-post responses to question 3 could indicate changes in knowledge about, or at least attitudes to, GM foods as a result of the lecture.

The results from this, or any, evaluation must be considered in context. This kind of evaluation offers some insight into the extent to which the short-term attitudes of audience members were moved as a result of their experience in the public lecture. If that was all you wanted to measure, then you have conducted a useful evaluation.

What such an evaluation will not reveal, however, is how enduring any measured changes will be or the extent to which any reported changes in attitude might lead to changes in behaviour with regard to GM food crops. If you were hoping to find out something about long-term change or effects on behaviour, then the goals and objectives that drove the evaluation were flawed, leading to inappropriate evaluation methods being employed.

## Summing Up

Efforts to evaluate a science communication, or indeed any, enterprise need not be complicated, time consuming or expensive, and the literature is literally teeming with excellent tools and resources, some of which are listed in Fig. 5.

**Tools and resources**

"A basic Guide to Program Evaluation" http://www.tgci.com/magazine/A%20Basic%20Guide%20to%20Program%20Evaluation.pdf

Community sustainability engagement: Evaluation Toolbox (accessed 10 September, 2012) http://evaluationtoolbox.net.au/index.php?option=com_content&view=article&id=11&Itemid=17

Foddy, W. (1999) *Constructing questions for interviews and questionnaires: theory and practice in social research*. Cambridge University Press: Melbourne

Papineau, D. and Kiely, M. (1996). Participatory evaluation in a community organization: Fostering stakeholder empowerment and utilization. *Evaluation and Program Planning*, 19 (1), 79-93.

Online survey generating tool - http://www.surveymonkey.com/

**Fig. 5** Evaluation tools and resources

Because of this abundance of existing evaluation resources, this chapter predominately focussed on providing critical perspectives from which to make decisions about how to choose an evaluation path.

Such decisions always begin with a clear articulation of the goal of the enterprise, or product, you wish to evaluate. From there, choices between evaluation tools, design, data gathering, analyses and subsequent reporting options are driven by broader contextual matters.

We also wanted to emphasise here the very real danger of tools becoming more important than their purpose. With an ever-increasing array of statistical packages and survey tools, and an environment where evaluations are regularly demanded but frequently not well understood, it is important to remember the fundamentals or you can easily become lost. This is why maintaining a mindful sense of perspective is essential.

The cautions and advice we offer here are perhaps best encapsulated in this quote from Donella Meadows:

> We try to measure what we value. We come to value what we measure. This feedback process is common, inevitable, useful, and full of pitfalls. (Meadows, 1998)

We hope this chapter will help you become more aware of some of these pitfalls and better still avoid at least some of them.

## References


Aitkin, D. (2012). *Ranking universities*. http://donaitkin.com/ranking-universities/. Accessed April 8, 2013.

Curry, S. (2012) *Sick of impact factors*. http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/. Accessed April 8, 2013.

Diekhoff, G. (1996). *Basic statistics for the social and behavioral sciences*. Upper Saddle River, NJ: Prentice Hall.

Donald, A. (2012). *Systematically wrong – From neutrinos to university league tables*. Occam's Corner. Science blog hosted by *The Guardian*. http://www.guardian.co.uk/science/occams-corner/2012/sep/06/systematically-wrong-neutrinos-university-league-tables?CMP=twt_fd. Accessed April 3, 2013.

Durant, J., Evans, G., & Thomas, G. (1989). The public understanding of science. *Nature, 340*, 11–14.

Falk, J. (2004). The director's cut: Toward an improved understanding of learning from museums. *Science Education, 88*, 83–96.

Fischer, J., Ritchie, E. G., & Hanspach, J. (2012). Academia's obsession with quantity. *Ecology & Evolution, 27*(9), 473–474.

Foddy, W. (1999). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Melbourne, Australia: Cambridge University Press.

Fowler, F. J., Jr. (2009). *Survey research methods* (Applied social research methods series, 4th ed.). L. Bickman & D. J. Rog (Series Eds.). Los Angeles: Sage.

Gibilisco, S. (2004). *Statistics demystified: A self-teaching guide*. Sydney, Australia: McGraw-Hill.

Lamberts, R., & Grant, W. J. (2012). *Australian R&D measures up globally . . . but what does that really mean?* https://theconversation.edu.au/australian-randd-measures-up-globally-but-what-does-that-really-mean-7085. Accessed March 15, 2013.

Meadows, D. (1998). *Indicators and information systems for sustainable development: A report to the Balaton Group*. Hartland Four Corners, VT: Sustainability Institute.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits. Committee on learning science in informal environments* (P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder, Eds.). Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

OECD. (2002). *Frascati manual* (6th ed.). Paris: OECD.

Oppenheimer, F. (1968). Rationale for a science museum. *Curator, 11*(3), 206–209.

Questacon. (2012). *2011–2012 annual review*. Canberra, Australia: Department of Industry, Innovation, Science, Research and Tertiary Education.

Rossi, P., Freeman, H., & Lipsey, M. (1999). *Evaluation. A systematic approach*. Thousand Oaks, CA: Sage.

Rumsey, D. (2003). *Statistics for dummies*. Hoboken, NJ: Wiley.

Sawyer, K. (2012). *Made to measure? Why university rankings are flawed*. https://theconversation.edu.au/made-to-measure-why-university-rankings-are-flawed-7306. Accessed February 15, 2013.

Şentürk E., & Özdemir, Ö. F. (2012). The effect of science centres on students' attitudes towards science. *International Journal of Science Education, Part B: Communication and Public Engagement*. Online. http://www.tandfonline.com/doi/full/10.1080/21548455.2012.726754

Shanghai Jiao Tong University Rankings. http://www.shanghairanking.com/. Accessed April 8, 2013.

Stocklmayer, S., & Bryant, C. (2011). Science and the public – What should people know? *International Journal of Science Education, Part B. Communication & Public Engagement*. Online http://dx.doi.org/10.1080/09500693.2010.543186

Tabachnick, B. S., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.

Vale, R. D. (2012). "Writing SMART objectives": Evaluating how we evaluate. *Molecular Biology of the Cell, 23*(17), 3285–3289.