How to — and How Not to — Evaluate Innovation

Burt Perrin

Evaluation 2002 8: 13
DOI: 10.1177/1358902002008001514

Evaluation
Copyright © 2002
SAGE Publications (London,
Thousand Oaks and New Delhi)
[1356–3890 (200201)8:1; 13–28; 021514]
Vol 8(1): 13–28

How to – and How Not to – Evaluate Innovation[1]

BURT PERRIN
Independent Consultant, France

Many traditional evaluation methods, including most performance measurement approaches, inhibit rather than support actual innovation. This article discusses the nature of innovation, identifies limitations of traditional evaluation approaches for assessing innovation and proposes an alternative model of evaluation consistent with the nature of innovation.

Most attempts at innovation, by definition, are risky and should 'fail' – otherwise they are using safe, rather than unknown or truly innovative approaches. A few key impacts by a minority of projects or participants may be much more meaningful than changes in mean (or average) scores. Yet the most common measure of programme impact is the mean. In contrast, this article suggests that evaluation of innovation should identify the minority of situations where real impact has occurred and the reasons for this. This is in keeping with the approach venture capitalists typically take where they expect most of their investments to 'fail', but to be compensated by major gains on just a few.

KEYWORDS: evaluation; innovation; learning; research and development; Research, Technology and Development (RTD)

The Nature of Innovation

Innovation can be defined as novel ways of doing things better or differently, often by quantum leaps versus incremental gains. This is consistent with the definition of innovation used by the European Commission's *Green Paper on Innovation* (1995: 1): 'the successful production, assimilation and exploitation of novelty in the economic and social spheres'. Innovation can be on a large scale, e.g. identification of a major new technology, a new business venture, or a new programme approach to a social problem. But it can also be on a small scale, involving initiatives within a larger project or programme, such as a teacher trying a new way of connecting with an individual student.

Innovation is sometimes used synonymously with the development or use of new technologies; but as the *Green Paper* indicates, the technological factor is just one potential element of innovation. One can be innovative in many other respects as well, e.g. better working conditions or methods of service delivery that may or may not have a technological component.

The above definition of innovation is consistent with concepts such as 'out of the box' thinking, double-loop learning (Argyris, 1982) and perhaps Drucker's (1998) concept of 'purposeful, focused change'.

By its very nature, innovation is:

- risky;
- unpredictable in terms of:
  - which particular activity or intervention will work or prove useful;
  - who will benefit;
  - when benefits, if any, will occur;
  - under which particular set of circumstances an innovative approach would be applicable;
  - whether the discovery and application will be as intended or of quite a different nature.

Hargadon and Sutton (2000), Al-Dabal (1998), Peters (1988), Zider (1998) and others have emphasized that 'success' comes from 'failure'. Innovation involves encouraging the generation of ideas and putting promising concepts to the test. One does not expect new concepts necessarily to work – indeed, if one is trying really new and unknown and hence, risky approaches, most should *not* work. As Zider (1998) has indicated:

> On average, good plans, people, and businesses succeed only one in ten times. . . . However, only 10% to 20% of the companies funded need to be real winners to achieve the targeted return rate of 25% to 30%. In fact, VC [venture capitalist] reputations are often built on one or two good investments. (p. 136)

A key component of effective innovation is an openness to learn. Drucker (1998), Hargadon and Sutton (2000), Khosla (Champion and Carr, 2000) and Peters (1988) emphasize that at least as much is learnt from 'failures' as from what *does* work. Drucker (1998) stresses that unexpected failure can be a major source of innovation opportunity, and that innovation most frequently works in ways different from expected. Peters (1988) says that many small failures can help avoid big failures. He suggests that one should 'Become a failure fanatic. Seek out little interesting foul-ups and reward them and learn from them. The more failure, the more success – period'. Khosla (Champion and Carr, 2000), who is considered one of the most successful of current venture capitalists, has indicated that:

> Our single biggest advantage may be the fact that we've screwed up in more ways than anybody else on the planet when it comes to bringing new technologies to market. That's a big institutional asset. Our hope is that we're smart enough not to repeat the old mistakes, just make new ones. (p. 98)

This approach is also consistent with the definition of learning, at least that of Don Michael, as Dana Meadows (2000) discusses.

> That's learning. Admitting uncertainty. Trying things. Making mistakes, ideally the small ones that come from failed experiments, rather than the huge ones that come

> from pretending you know what you're doing. Learning means staying open to experiments that might not work. It means seeking and using – and sharing – information about what went wrong with what you hoped would go right.

The above concept of innovation is also consistent with Donald T. Campbell's theory of evolutionary epistemology (e.g. Campbell, 1974, 1988a; also see Shadish et al., 1991), based upon the Darwinian metaphor of natural selection. He claims that a blind-variation-and-selective-retention process is a basic component of all genuine increases in knowledge, involving these three critical mechanisms:

1. Generation of a wide range of novel potential solutions;
2. Consistent selection processes; and
3. A means of preserving the selected variations.

Campbell has emphasized the importance of trial and error, and in particular trying out a wide range of bold potential 'variants', including approaches that may seem unlikely to work, provided that these are subject to evaluation. This, for example, is consistent with his view of an 'experimenting society' (Campbell, 1969, 1971, 1988b). It may also be consistent with the approach promoted by the Blair Government in the United Kingdom of using evaluation, in particular of pilot projects, to provide 'evidence-based policy' (although a lively topic of debate at recent UK Evaluation Society conferences has been the extent to which short-term pilots provide sufficient time for meaningful evaluation of significant social reforms [also see Martin and Sanderson (1999) on this topic]).

Innovations are generally long term in nature, sometimes very long term. As Drucker (1998) indicates, the progress of innovation is uneven rather than continuous and the payoff is rarely immediate. One cannot do meaningful evaluation of impact prematurely; attempting to assess 'results' too soon can be counterproductive to the innovative process. As Drucker (1998: 156) has indicated 'knowledge-based innovations [have] the longest lead time of all innovations. . . . Overall, the lead time involved is something like 50 years, a figure that has not shortened appreciably throughout history'. Georghiou (1998) similarly indicates that it can take considerable time for project effects to become evident, e.g. referring to a Norwegian study indicating that some 12–15 years are needed for outcomes to become clear.

Buderi (2000) indicates that corporate research today is looking, mainly, for shorter-term payback. Nevertheless, this is not expected to be instant or on command. Businesses expect a variety of different levels of innovation ranging from short-term, minor fine-tuning over a 1–2-year period, to the development of new products over an intermediate period, to the generation of revolutionary ideas that completely change the nature and business of the organization and are essential for long-term survival.

Even though innovation, by definition, is risky and deals with the unknown, it does not follow that it is facilitated in a laissez-faire manner. For example, the notion of calculated risk is basic to venture capitalists, who (generally) do extensive analysis before making any investment, even though they expect to win on only a select few. It is generally recognized that while it is challenging, it is nevertheless critical to manage innovation. The National Audit Office in the UK (2000), in a recent report, emphasizes the importance of managing risk. It is increasingly recognized that even fundamental research needs to be linked in some way with potential users and applications (e.g. Buderi, 2000). This, and implications for evaluation, are discussed in more detail below.

Limitations of Typical Approaches to Evaluation

Inappropriate Use of Mean Scores to Assess Impact

Evaluation conclusions are most commonly based upon mean (or average) scores:

- The basis of the experimental and quasi-experimental approach is to compare the mean of the experimental group with that of the control or comparison group.
- There is an implicit assumption in quantitative data gathering and analysis that more is invariably better, e.g. a rating of 67 percent improved is usually considered positively, while if 'only' 20 percent show a benefit, this generally is not.
- Most evaluations look for the percent 'success' rate: the numbers and proportion of participants who have succeeded or benefited on some criterion or other. They implicitly or explicitly fail to acknowledge that just a few 'successes' can make a programme worthwhile. Evaluation approaches that look for the percent of 'success' do not acknowledge that with innovation one invariably succeeds via the small number of exceptions, and usually after a series of 'failures'.

For example, a funding programme may have just a 1-percent 'success' rate, but if one project out of 100 results in a cure for AIDs, surely the funding of the other 99 attempts does not represent a 'failure'. This may appear to be obvious, but the same can apply to programmes attempting to find innovative solutions to youth unemployment, rural poverty, pesticide reduction, etc. where a low percentage of 'successful' projects would most likely be seen as a problem.

Mean scores invariably hide the true meaning and the most important findings. For example, one can obtain a mean rating of 3 out of 5 when *all* respondents achieve a score of 3. But one can equally achieve a mean of 3 when *none* of the respondents get this rating, for example when half have a score of 1 and the other half a score of 5. These hypothetical situations represent radically different outcomes, which nevertheless are hidden if one just looks at the mean. Yet it is not uncommon to see research reports, even those issued by evaluation departments of well-respected international organizations, to present mean scores without showing any breakdowns or distributions or measures of variance.

Consider a real-life example of how mean scores can disguise what is actually taking place. As Perrin (1998) discusses, the median household income for US households in 1996 was reported to have increased 1.2 percent over the previous year. One gets a very different picture, however, if this figure is broken down by wealth, where income of the wealthiest 20 percent increased by 2.2 percent and that of the middle 60 percent by 1.1 percent – but the income of the poorest 20 percent decreased by 1.8 percent.

Simplistic Models of Impact

Smith (2000) emphasizes the importance of a systems perspective with respect to innovation and knowledge creation, given that innovation never occurs alone but always within a context of structured relationships, networks, infrastructures and in a wider social and economic context. He indicates that an interactive model of innovation has emerged, that 'linear notions of innovation have been superseded by models which stress interactions between heterogeneous elements of innovation processes' (p. 16). Similarly, the European Commission's MEANS evaluation guidelines indicate that 'The linear "Science-Technology-Production" type model has given way to the conceptualization of innovation as a dynamic, interactive and non-linear process' (European Commission, 1999: 31).

Nevertheless, in Europe and elsewhere, there is still considerable evaluation activity that assumes a direct relationship between input and output, including many evaluations that attempt to specify the return on investment in science and other forms of innovation. As Georghiou (1998) discusses, this approach inappropriately assumes a direct cause–effect relationship. Jordan and Streit (2000), Branscomb (1999) and others also discuss the limitations of this and similar models, and the need for a new conceptual model for discussing and evaluating public science. This model needs to acknowledge that the nature of the impact of innovation is mediated through context and interaction with many other activities.

Campbell (1974, 1988a) indicated that mechanisms for innovation or generation and for preservation or retention are inherently at odds. Evaluation approaches drawn from frameworks that assume the preservation of the status quo are likely to apply criteria inappropriate for assessing programmes and approaches that seek innovative alternatives. Davies (1995), House (2000) and Stronach (2000a, 2000b) have, respectively, described the devastating consequences of inappropriate use of traditional evaluation models for assessing: development programmes in Bangladesh; an innovative education programme aimed at high-risk black youth in Chicago, sponsored by the Reverend Jesse Jackson; and the non-traditional Summerhill School in the UK that the OFSTED (official government) inspection process initially had found wanting and had recommended closing.

Misuse of Performance Measurement Approaches

Performance measurement is increasingly being used as a means of evaluation of research, technology and development (RTD) and other initiatives presumably based upon innovation (e.g. Georghiou, 1998; Jordan and Streit, 2000). Performance indicator or objective-based approaches to evaluation can be useful for monitoring purposes, in particular for tracking project status to ensure that innovative activities are more or less on track. Arundel (2000) suggests that indicators (or 'innovation scorecards') can be useful at a macro level, e.g. in building consensus about the need for policy action in support of research. He adds, however, that they are not relevant at the meso and micro levels, where most activities and most policy actions occur.

More to the point, performance measures or indicators are rarely appropriate for assessing impact. Given that innovation by definition is unpredictable, it is not possible to identify meaningful objectives or targets in advance. Evaluation approaches largely based upon assessing the extent to which programmes have achieved pre-determined objectives ipso facto are not open to double-loop learning, and can penalize programmes that go beyond or demonstrate limitations in these objectives. Furthermore, true gains, including the identification of what can be learned from 'failures' as well as from 'successes', can be difficult or impossible to quantify. As Blalock (1999), Davies (1999), Greene (1999), Mintzberg (1996), Perrin (1998) and others have pointed out, performance indicators and evaluation-by-objectives by themselves are rarely suitable for evaluating any programme, innovative in intent or not. Smith (2000) adds that recent developments in theories of technological change have outrun the ability of available statistical material to be relevant or valid.

Nevertheless, I have seen research funding programmes required to report 'results' – in quantitative performance indicator terms – on a quarterly basis! The result is a strong disincentive to innovative activity and a bias towards that which is short-term and less risky. In order to meet performance targets with any certainty, it would only make sense to fund research to explore what is already known; to do otherwise would be too risky. This can apply equally to innovation in areas other than research, including almost any area of public policy. Thus the perverse consequence of performance measurement can be less, rather than more innovation and true impact.

The Reactive Nature of Evaluation Resulting in Less Innovation

Many traditional approaches to the evaluation of innovation fail to recognize the reactive nature of evaluation. Just as performance indicators reward safe, short-term activities, evaluations based upon mean scores rather than upon the recognition of the few but extraordinary accomplishments punish innovation and those who explore the unknown. Instead, they reward mediocrity. The unintended result is to discourage people from trying anything truly innovative. 'Failures' are usually viewed and treated negatively, with negative consequences for those judged to have 'failed', even if the attempt was very ambitious.

Indeed, any programme or project claiming to be innovative that has a high record of 'success' should be viewed with scepticism as this probably means that what is being attempted is not very ambitious. The result is more likely to be mediocrity, in contrast with programmes that have a high number of 'failures'.

Similarly, funding programmes themselves tend to be viewed as failures if most of the projects or activities they fund are not 'successful' or run into major problems. This may occur even if there are major advances in a few of the projects and much has been learnt from many of the others. Again, the unintended consequence of evaluations that take this attitude is less – rather than more – innovation.

The National Audit Office (2000) in the UK, in its report promoting risk management, encourages civil servants to innovate, take risks and move away from a blame culture. However, despite this laudable intent, it seems likely that risk management, the recommended approach, will instead be interpreted as the necessity to minimize risk and the requirement to have a solid paper trail documented in advance to justify anything that turns out to be 'not fully successful'. The danger that this approach, despite its intention, instead will result in the inhibition of innovation has been identified in an Annex to the report's executive summary, independently prepared by Hood and Rothstein, who indicate that:

> Risk management if inappropriately applied can serve as a fig-leaf for policy inaction . . . or as an excuse for sticking to procedural rules . . . [and] would also further obstruct processes of learning from mistakes. (Hood and Rothstein, 2000: 27)

The disincentive to true innovation can be very real. For example, I have spoken to a number of people and representatives of projects who have indicated their fear even of using funds designated specifically for innovative activities to attempt the unknown for fear of what would happen if they did not succeed. They say that if they really want to try something risky, they need to do so outside the parameters of the funding programme.

Similarly, I have spoken with officials with research funding programmes in the European Commission and in Australia who have acknowledged that despite the brief for their programmes, they are 'not very innovative'. Instead, they are forced to fund mainly safe projects, for fear of the consequences of 'failure'.

As a result, many true innovations come from 'fugitive' activities, or from those brave individuals who dare to push the limits and brave the consequences.

Alternative or Innovative Approaches to the Evaluation of Innovation

So, how *should* one approach the evaluation of projects, activities or programmes that are or should be considered innovative? Following are some suggestions.

Take a Key-Exceptions or Best-Practices Approach to Evaluation

When evaluating innovation, one should use criteria similar to those employed by venture capitalists in assessing the value of their investments. They expect that only a small minority of their investments will, eventually, strike it big. It is considered a learning opportunity rather than a problem that as many as 80–90 percent of their investments do not work out well, or even collapse completely. Similarly, evaluators could put greater emphasis on identifying positive examples ('best practices'), rather than 'averages', even if they are small in number; additionally, other lessons that might arise from 'failures' as much as from 'successes' could be emphasized.

Language should be used, both overtly and implicitly, with care in data interpretation and reporting. For example, one should be careful of making statements such as '"only" 10 percent of funded projects demonstrated positive results'. If 'just' one out of 20 projects exploring innovative ways of, e.g. training unemployed people, or addressing rural poverty demonstrates positive results, and does so in a way that can inform future practice, then the programme has accomplished something very real. This surely can be a more meaningful finding than if most projects demonstrate marginal positive gains. Similarly, if 'only' two out of 20 demonstration projects 'work', this is not necessarily a negative finding, particularly if implications for future directions can be identified.

Use a Systems Model

As discussed earlier, the innovative process is not linear. Innovations rarely come from 'lone wolf' geniuses working alone, but instead through partnerships and joint activities within a much wider social and economic context. Outcomes resulting from the application of innovative ideas almost always arise from interaction with many other factors. As Jordan and Streit (2000) and many others emphasize, innovation is only one factor contributing to the effectiveness of science and technology organizations. A simple input–output or cause-and-effect model of evaluation is not appropriate.

Consequently, it would seem that a systems approach that considers the workings of an innovative approach may be applicable in many instances. A systems approach has the potential, as Smith (2000) has indicated, of being able to explore the *dynamics* of the innovation and knowledge creation process. These dynamics and interactions may be more important than any single intervention. In particular, this approach would appear particularly appropriate when looking at large-scale innovations, such as those at an organizational level, as well as others cutting across multiple organizations or at a societal level. Nevertheless, there appear to be limited examples of the effective use of systems approaches in evaluation. This seems to be an area where more attention would be warranted.

Focus on Learning and the Degree of Innovation, rather than 'Successes'

Evaluations of innovative projects and programmes should identify the extent to which there has been any attempt:

- to learn from 'failures' (as well as from 'successes');
- to identify implications for the future; and
- the extent to which action has been taken based upon what has been learned.

A learning approach to the evaluation of innovation can be more important than tabulating the number of successful 'hits'. Particularly at the programme or funding level, evaluation should focus on the extent to which findings and implications have been identified and disseminated – based upon the funding agency's own practices as well as the activities of its funded projects. Some funding programmes of innovative approaches are very good at this. Others, including some funding programmes with explicit objectives stating their own openness to learn from 'failures' as well as from 'successes', seem to either do it imperfectly, or not at all. Evaluation can play a useful role by pointing this out. Box 1 suggests some criteria which may be used for the evaluation of agencies with a mandate to support innovation.

As indicated earlier, one can learn at least as much from what has *not* worked as from what has. Evaluation should also recognize that 'failure' may represent work in progress. Additionally, it should be noted that progress, especially as a

Box 1. Suggested Criteria for the Evaluation of Agencies or Programmes Supporting Innovation

- How ambitious is the funding agency, with respect to its own practices as well as with the projects and activities that it funds?
- Do a significant proportion of funded activities 'fail'? (If not, is there any other evidence demonstrating that these were actually innovative?)
- How soon does the agency 'pull the plug' on projects that have not (yet) demonstrated 'success'?
- Does the agency identify what has been learned from its funded projects and implications for future directions – including from those that have 'failed' as well as from the 'successes'? To what extent does it attempt to synthesize key findings and implications from across specific settings or projects, using a cluster evaluation approach or other means of synthesis?
- Are lessons learned and implications disseminated, in appropriate language for the intended audience (i.e. in other than technical language for those with the potential to use or apply the information)?
- Does action of some form follow from what has been learned?
- To what extent does the agency stimulate, support, and reward risk taking, both internally amongst its staff and externally amongst its funded projects and key constituencies?
- To what extent is the agency calculated in its risk taking, e.g. through consideration of how much risk is appropriate or not, distinguishing between risky and poor quality proposals, identification of where there seems to be the greatest potential for major learning, or through other forms of risk management?
- Is some proportion of staff time and funding set aside to pursue ideas that do not fit into any established categories?
- How does the agency manage innovation and monitor and support the projects that it funds?

result of significant innovation, is uneven, and generally occurs in quantum leaps after a long period of uncertainty, rather than as incremental gains.

'Success' and 'failure', of course, are not dichotomous, but endpoints on a multi-dimensional continuum. There can be degrees both of 'success' and of 'failure', as well as differences of opinion about how the performance of a given initiative should be classified; this is especially true when there is a lack of clear goals, which is commonplace with many social programmes. Even the most successful programmes can, and should, have tried various techniques that may not have worked out perfectly. The fact that a programme continues to exist, be it a private sector business or a social programme, does not mean that it is necessarily 'successful' (or will continue to be so) and cannot be improved. With a learning approach that emphasizes what one can do to improve future effectiveness, there is less need to make summative judgements about 'success' or 'failure'.

Evaluation itself can play a major supportive role in helping to identify lessons learned and implications for future directions. Indeed, this can represent a major reason to undertake evaluation of innovative programmes. Along these lines, there may be opportunities for greater use of cluster evaluation approaches (e.g. Perrin, 1999; Sanders, 1997). There also appears to be a greater need for identifying and disseminating information about what has *not* worked, as well as the 'successes', to help avoid repeated 'reinvention of the square wheel'.[2]

As a corollary, another major criterion for evaluation should be the degree of ambitiousness or innovation of what was attempted. Projects and activities that are truly ambitious in nature, breaking new limits and trying out new ideas, should be recognized and rewarded, whether or not they have 'worked' as intended. The criteria for success should not be whether the project succeeded or failed in what it was trying to do, but rather should be the extent to which it truly explored something new, identified what can be learned and acted upon these. This is consistent with Elliot Stern's (1999) recommendations to a parliamentary committee.

Set Realistic Timeframes

As discussed earlier, major innovations rarely can be developed or properly assessed in the short term. Certainly three months (I have seen this occur) or 12 months (the most common timeframe) is much too soon to evaluate the impact of most innovative activities. For example, frequently there is a tendency to evaluate the impact of pilot or demonstration projects before they have had a chance to get established and to work through the inevitable start-up problems. Logic models can help in identifying what forms of impact are appropriate to look for at given stages in a project cycle. While practical constraints dictate undertaking evaluation at an early stage, one should be explicit about this and be very cautious in drawing conclusions about impact.

This problem has been acknowledged, at least in part, by DG Research of the European Commission (e.g. see Airaghi et al., 1999). Evaluation of its Fourth European RTD Framework Programme is continuing even after implementation of the Fifth Framework Programme. (Of course, many of the funded research projects are multi-year in nature, and still in process at the conclusion of the funding programme.)

Incorporate a Process Approach

Evaluation of innovation may take a process approach, identifying the extent to which projects embody those characteristics or principles known to be associated with innovation and the values or goals of the sponsoring agency. Perhaps a related evaluation question might be the extent to which innovation is being managed, e.g. to encourage the identification and application of innovative ideas and approaches.

The specific principles or characterisics that should be considered depend upon the particular topic area being evaluated. For example, venture capitalists typically consider criteria such as: the extent to which a company has sufficient capital; capable and focused management; a good idea with market potential; skilled and committed staff, etc. Criteria I have used for assessment of research include:

- legitimacy (e.g. appropriateness and priority of the research as assessed by user groups);
- potential relevance and application;
- quality of the research;
- contact with or involvement of potential user groups;
- identification of what has been learned, potential applications and implications;
- dissemination of findings and implications to a range of audiences, in particular to potential users as well as other researchers, in non-technical language;
- extent of partnership and collaboration;
- extent to which the idea is new or different, truly innovative;
- openness of the research approach to serendipity and unexpected findings.

The above list draws in part upon an increasing literature (e.g. Buderi, 2000; Jordan and Streit, 2000; Kanter, 1988; Zakonyi, 1994a, 1994b) indicating characteristics of organizational culture and environment which appear to be most closely associated with the presence of innovation at various stages. Buderi in particular emphasizes how ongoing contact and involvement between the researcher and potential users play a key role in enhancing the value of innovation. To a large extent, compliance with the above and with similar sets of principles can be assessed ex ante, as well as concurrently and ex post.

Thus innovation in research, even fundamental research, is tied to consideration of potential relevance, close contact with potential users, and attempts to identify applications. The corporate research world has moved away from carte blanche research. Nevertheless, leading corporate research organizations typically leave some portion of research budget and researcher time for projects that do not fit into established categories. For example, often up to 25 percent of the research budget is left open to ideas that do not conform to existing categories (e.g. Buderi, 2000). The European Commission is considering a similar approach to provide for funding of 'blue sky' research proposals. 3M is an example of a corporation, known for its innovation, that lets its researchers devote 10 percent of their time to activities of their own choosing (Shaw et al., 1998).

This also means that some typical approaches to the evaluation of research, e.g. numbers of publications, presentations, scientific awards, or peer or 'expert' assessments of research quality, etc. are irrelevant and inappropriate. Nevertheless, as Georghiou (1998) and others have indicated, these approaches, for the evaluation of research institutions in particular, are still commonplace.

Use Appropriate Methodologies

A methodological approach to the evaluation of innovation must be able to:

- get at the exceptions, including unintended consequences, given that research approaches just based upon counting and summations are not relevant and will hide true achievements;

- provide an understanding of the complex processes involved as well as help identify learning and implications, from 'successes' and 'failures';
- be flexible enough to be open to serendipity and unexpected findings, which, particularly with innovations, can represent the key outcomes.

Qualitative methods, by themselves or possibly in combination with other approaches, are particularly suitable for questions such as the above (e.g. Patton, 1990). Case study designs would seem especially applicable (e.g. Yin, 1994). This would permit exploration in detail of both apparent 'successes' and 'failures', to identify what it is that does or does not make them work and what can be learned in either case. When the primary focus is on learning, intentional rather than random sampling may be most appropriate.

Quantitative methods may be appropriate, provided that they are not used in isolation. For example, quantitative analysis could be used to suggest where to focus detailed 'qualitative' attention on potentially intriguing findings. One should be cautious, however, when using quantitative data for assessing innovation. They should be used only where meaningful, not just because they are easier to attain and to count than qualitative data.

When carrying out quantitative analysis, one should be cautious about aggregation, using mean scores as starting points to ask questions of the data and for further exploration, such as why some projects or activities seem to be working differently than others. In particular, the data should be broken down and the variations and outliers assessed, recognizing that impact with respect to innovation comes mainly from the outliers or exceptions (e.g. see Miles and Huberman, 1994).

Any form of evaluation methodology can be appropriate to assess the impact of a given project, to determine if there has been an innovative discovery and application. The appropriate choice of methodology will depend upon the particular type of project or activity, the evaluation questions and other factors.

Acknowledge Political and Organizational Realities

Weiss (e.g. 1999, 2000) has emphasized how the policy and decision-making process is anything but rational, where 'objective' research represents just one consideration among many involving multiple competing actors and interests. For example, politicians faced with intense pressure to act quickly, even in the absence of evidence indicating a clear course of action, are notorious for their short-term outlook and for limited interest in impact that may not occur until some time after their own tenure in office (but the same pressures may also apply in the private sector, where addressing expectations of the investment community on the next quarterly report may take priority over long-term considerations).

There is often strong pressure within the public sector to focus more on avoiding mistakes than on attempting risky approaches that may or may not work as expected. For example, the UK National Audit Office (2000) emphasizes the need to break the 'culture of blame' that is too often pervasive within public services. As Shadish et al. (1991) observed, Campbell was pessimistic about the extent of true innovation that most governments offer even in the name of 'reforms' purportedly intended to address real problems and to achieve social change. Maddy (2000) indicates that while private sources of funding recognize that high reward is accompanied by high risk, government and quasi-government organizations:

> . . . are terrified of risk and deeply enmeshed in bureaucracy and their own rigid methods of investment and analysis. They are not necessarily looking for big paybacks on their investments. They are more preoccupied with adhering to their established procedures. (p. 64)

Yet there is increasing recognition and talk of the need for more innovation within the public sector (e.g. European Commission, 1995; National Audit Office, 2000). To achieve this it will be necessary at least to recognize disincentives and attempt to address these. Changing organizational culture would involve inevitable compromises, but it is not impossible to provide greater incentives and opportunities for innovative approaches that would involve some degree of risk.

The identification and open acknowledgement of factors impeding innovation would be a good way to begin, followed by the consideration of how these factors could be addressed. For example, the National Audit Office (2000) has indicated that the system of rewards and punishments needs to be changed. One might look for opportunities to publicly reward managers and staff who have attempted to innovate in some way, even if their initiatives have *not* been as 'successful' as has been hoped. Another approach that some organizations have taken is to establish special funds where 'blue sky', or risky approaches, that otherwise would not fit into other categories, can be supported.

The approach to evaluation of innovation can also play a key role. Many approaches to evaluation can, perhaps unintentionally, act as disincentives to innovation. Conversely, evaluation approaches that recognize rather than punish ambitiousness and identify what can be learned from what has been tried, irrespective of outcome, can play a significant role in supporting a culture of innovation.

Conclusion

Most attempts at innovation, by definition, *must* fail. Otherwise, they are not truly innovative or exploring the unknown. However, value comes from that small proportion of activities that are able to make significant breakthroughs, as well as from identifying what can be learned from 'failures'.

When evaluating innovation, it is important to note how mean or average scores can mislead and disguise what is truly happening. It is important to remember that evaluation is reactive. If it punishes those who try something different, or is viewed in this light, it can act as a disincentive to innovation. In contrast, evaluation can be invaluable in helping to identify what can be learned both from 'successes' and 'failures' and implications for future directions. There may be opportunities to be more innovative about how we evaluate innovation.

Notes

1. Earlier versions of this paper were presented at the UK Evaluation Society conference, London, 8 December 2000, as well as at the European Evaluation Society conference, Lausanne, 13 October 2000.
2. Analogy identified in conversation with Mel Mark, October, 2000.

References

Airaghi, A., N. E. Busch, L. Georghiou, S. Kuhlmann, M. J. Ledoux, A. F. J. van Rann and J. V. Baptista (1999) *Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes*. Report by the Independent Reflection Group of the European Technology Assessment Network (ETAN) to the European Commission DG XII, Evaluation Unit. Brussels.

Al-Dabal, J. K. (1998) 'Entrepreneurship: Fail, Learn, Move On', unpublished, paper, Management Development Centre International, The University of Hull.

Argyris, C. (1982) *Reasoning, Learning, and Action*. San Francisco, CA: Jossey-Bass.

Arundel, A. (2000) 'Innovation Scoreboards: Promises, Pitfalls and Policy Applications', paper presented at the Conference on Innovation and Enterprise Creation: Statistics and Indicators, France, 23–4 November.

Blalock, A. B. (1999) 'Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration', *Evaluation* 5(2): 117–49.

Branscomb, L. M. (1999) 'The False Dichotomy: Scientific Creativity and Utility', *Issues in Science and Technology* 16(1): 6–72.

Buderi, R. (2000) *Engines of Tomorrow: How the World's Best Companies are Using Their Research Labs to Win the Future*. London: Simon & Schuster.

Campbell, D. T. (1969) 'Reforms as Experiments', *American Psychologist* 24: 409–29.

Campbell, D. T. (1971) 'Methods for the Experimenting Society', paper presented at the Meeting of the Eastern Psychological Association, New York, and at the Meeting of the American Psychological Association, Washington, DC.

Campbell, D. T. (1974) 'Evolutionary Epistemology', in P. A. Schilpp (ed.) *The Philosophy of Karl Popper*. La Salle, IL: Open Court. Reprinted in D. T. Campbell (E. S. Overman, ed.) (1988a) *Methodology and Epistemology for Social Science: Selected Papers*. Chicago and London: University of Chicago Press.

Campbell, D. T. (E. S. Overman, ed.) (1988) *Methodology and Epistemology for Social Science: Selected Papers*. Chicago, IL and London: University of Chicago Press.

Campbell, D. T. (1988b) 'The Experimenting Society', in E. S. Overman (ed.) *Methodology and Epistemology for Social Science: Selected Papers*. Chicago, IL and London: University of Chicago Press.

Champion, D. and N. G. Carr (2000) 'Starting Up in High Gear: An Interview with Venture Capitalist Vinod Khosla', *Harvard Business Review* 78(4): 93–100.

Davies, I. C. (1999) 'Evaluation and Performance Management in Government', *Evaluation* 5(2): 150–9.

Davies, R. (1995) 'The Management of Diversity in NGO Development Programmes', paper presented at the Development Studies Association Conference, Dublin, September. (Available online at: http://www.swan.ac.uk/eds/cds/rd/diversity.htm)

Drucker, P. F. (1998) 'The Discipline of Innovation', *Harvard Business Review* 76(6): 149–56.

European Commission (1995) *Green Paper on Innovation*. (Available online at http://europa.eu.int/en/record/green/gp002en.doc)

European Commission (1999) *MEANS Collection – Evaluation of Socio-Economic*

Programmes. Vol. 5: Transversal Evaluation of Impacts in the Environment, Employment and Other Intervention Priorities.

Georghiou, L. (1998) 'Issues in the Evaluation of Innovation and Technology Policy', *Evaluation* 4(1): 37–51.

Greene, J. C. (1999) 'The Inequality of Performance Measurements', *Evaluation* 5(2): 160–72.

Hargadon, A. and R. I. Sutton (2000) 'Building an Innovation Factory', *Harvard Business Review* 78(3): 157–66.

Hood, C. and H. Rothstein (2000) 'Business Risk Management in Government: Pitfalls and Possibilities', in National Audit Office *Supporting Innovation: Managing Risk in Government Departments*, Annex 2, Report by the Comptroller and Auditor General. HC864 1999/2000, pp. 21–32. London: The Stationery Office.

House, E. R. (2000) 'Evaluating Programmes: Causation, Values, Politics', keynote address at the UK Evaluation Society Annual Conference, London, 7 December.

Jordan, G. B. and L. D. Streit (2000) 'Recognizing the Competing Values in Science and Technology Organizations: Implications for Evaluation', paper presented at the US/European Workshop on Learning from Science and Technology Policy Evaluation, Germany, September.

Kanter, R. M. (1988) 'When a Thousand Flowers Bloom: Structural, Collective and Social Conditions for Innovation in Organizations', *Research in Organizational Behavior* 10: 169–211.

Maddy, M. (2000) 'Dream Deferred: The Story of a High-Tech Entrepreneur in a Low-Tech World', *Harvard Business Review* 78(3): 57–69.

Martin, S. and I. Sanderson (1999) 'Evaluating Public Policy Experiments: Measuring Outcome, Monitoring Processes or Managing Pilots?', *Evaluation* 5(3): 245–58.

Meadows, D. (2000) 'A Message to New Leaders from a Fallen Giant', *The Global Citizen*. 9 Nov. (Also available online at http://www.sustainer.org)

Miles, M. B. and A. M. Huberman (1994) *Qualitative Data Analysis*. Thousand Oaks, CA and London: Sage.

Mintzberg, H. (1996) 'Managing Government, Governing Management', *Harvard Business Review* 74(3): 75–83.

National Audit Office (2000) *Supporting Innovation: Managing Risk in Government Departments,* Report by the Comptroller and Auditor General, HC864 1999/2000. London: The Stationery Office.

Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods*, 2nd edn. Thousand Oaks, CA and London: Sage.

Perrin, B. (1998) 'Effective Use and Misuse of Performance Measurement', *American Journal of Evaluation* 19(3): 367–79.

Perrin, B. (1999) 'Evaluation Synthesis: An Approach to Enhancing the Relevance and Use of Evaluation for Policy Making', presentation to the UK Evaluation Society Annual Conference, Edinburgh, December.

Peters, T. (1988) *Thriving on Chaos: Handbook for a Management Revolution*. London: Pan Books.

Sanders, J. R. (1997) 'Cluster Evaluation', in E. Chelimsky and W. R. Shadish (eds) *Evaluation for the 21st Century*, pp. 396–404. Thousand Oaks, CA and London: Sage.

Shadish, W. R., T. D. Cook and L. C. Leviton (1991) *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks, CA and London: Sage.

Shaw, G., R. Brown and P. Bromiley (1998) 'Strategic Stories: How 3M is Rewriting Business Planning', *Harvard Business Review* 76(3): 41–50.

Smith, K. (2000) 'Innovation Indicators and the Knowledge Economy: Concepts, Results and Policy Challenges', keynote address at the Conference on Innovation and Enterprise Creation: Statistics and Indicators, France, 23–4 November.

Stern, E. (1999) 'Why Parliament Should Take Evaluation Seriously', *The Evaluator*, November.

Stronach, I. M. (2000a) 'Expert Witness Statement of Ian MacDonald Stronach', to the Independent Schools Tribunal in the case between Zoe Redhead (Appellant) and the Secretary of State for Education and Employment (Respondent), 21 February.

Stronach, I. M. (2000b) 'Evaluating the OFSTED Inspection of Summerhill School: Case Court and Critique', presentation to the UK Evaluation Society Annual Conference, London, 7 December.

Weiss, C. H. (1999) 'The Interface between Evaluation and Public Policy', *Evaluation* 5(4): 468–86.

Weiss, C. H. (2000) 'The Experimenting Society in a Political World', in L. Bickman (ed.) *Validity and Social Experimentation: Donald Campbell's Legacy*, Vol. 1, pp. 283–302. Thousand Oaks, CA and London: Sage.

Yin, R. K. (1994) *Case Study Research: Design and Methods*, 2nd edn. Thousand Oaks, CA and London: Sage.

Zakonyi, R. (1994a) 'Measuring R&D Effectiveness I', *Research Technology Management* 37(2): 27–32.

Zakonyi, R. (1994b) 'Measuring R&D Effectiveness II', *Research Technology Management* 37(3): 44–55.

Zider, B. (1998) 'How Venture Capital Works. The Discipline of Innovation', *Harvard Business Review* 76(6): 131–9.

BURT PERRIN is an independent consultant now based in France. He consults around the world to international organizations, governments, and non-governmental private organizations in the areas of evaluation, applied research, organizational learning and development, strategic planning and policy development, and training. He takes a practical approach to his work, striving to help develop capacity and expertise in others. Please address correspondence to: La Masque, 30770 Vissec, France. [email: burt_perrin@compuserve.com]