Summary of Public Sector Guidance for Evaluation with respect to STEM Initiatives

Issue

1. This document aims to establish a joint set of guidelines for the Department for Children School and Families (DCSF) and the Department for Innovation Universities and Skills (DIUS) to implement when undertaking evaluations of STEM initiatives, with a view to encouraging comparability across evaluation of initiatives. It is also a useful guide to other organisations with responsibility for STEM initiatives looking to carry out evaluation.

2. Initiatives are defined widely in this guidance. Depending on the context, it refers to both specific activities and programmes of activities, as well as the organisations that carry out such activities.

Summary

3. This document relies on Her Majesty's Treasury Green Book (and DCSF/DIUS) guidance on evaluation procedure to set out a standard practice for evaluating STEM initiatives. Similar guidance is available on the DIUS intranet. It also draws on advice from the Research Councils UK publication 'Evaluation: Practical Guidelines'[1] which offers advice on evaluation for those in the STEM field. It covers four main topic areas, including:
    a. Introduction to Evaluation
    b. What is to be evaluated
    c. How is it to be evaluated
    d. Feedback and publication

4. Each of these broad headings will contain guidelines for the standard approach for undertaking an evaluation in the area of STEM.

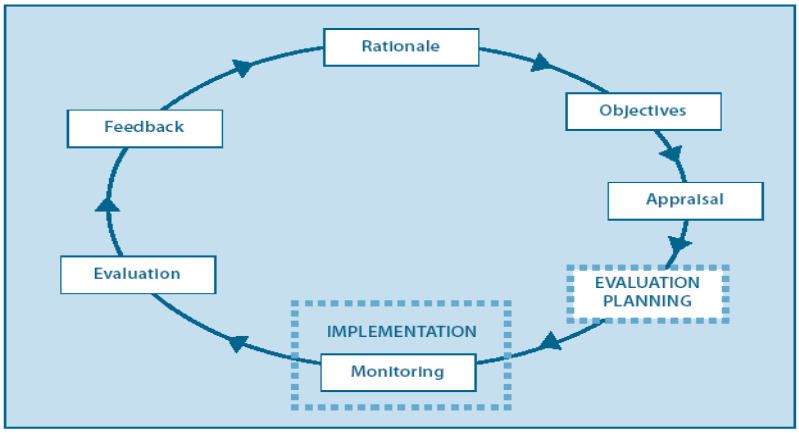
Introduction to Evaluation

5. An evaluation should be an integral part of initiative design; it forms one part of a cycle of implementation of an initiative. The cycle of initiative design is known as ROAMEF (Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback) and is discussed in more detail in the DCSF guidance; the diagram below outlines the key process.

---

[1] Green Book:
http://www.hm-treasury.gov.uk/economic_data_and_tools/greenbook/data_greenbook_index.cfm
DCSF/DIUS Evaluation and Appraisal Guidance:
http://www.dfes.gov.uk/rsgateway/DB/ECO/e000999/EvalAppGuide2.pdf
RCUK Evaluation Guidelines:
http://www.rcuk.ac.uk/aboutrcuk/publications/corporate/evaluationguide.htm

Figure 1: The 'ROAMEF' Policy Making Cycle



6. Evaluation results inform decisions as to whether existing policies or strategies should be continued or modified. On a longer-term horizon, they provide lessons to improve subsequent rounds of expenditure and policy making. Appraisal is the assessment of the likely impacts of an initiative before it is implemented. It is commonly used to assess the preferred choice between different options. Evaluation is very similar to appraisal, but rather than being based on informed predictions and forecasts like appraisal, evaluation is carried out once we know what actually happened. To effectively conduct evaluation, it should be thought about and planned **before** an initiative is implemented.

7. As is clear from the ROAMEF diagram monitoring is part of evaluation. In most cases the information collected to monitor the initiative will form the basis of the evaluation.

8. In planning an evaluation, adequate resource should be devoted to managing the evaluation and assimilating the results it produces. The value of the information contained in an evaluation is only realised when it is understood and used to develop further interventions.

What is to be evaluated?

9. Vital in any evaluation is clearly specifying the activity that is to be evaluated; this ensures that data collection is focused on the right areas. This involves defining and quantifying:
    a. Objectives/Targets – *what the initiative is trying to achieve*
    b. Inputs – *the processes or actions put into place to implement the initiative*
    c. Outputs – *the measures which can be used to evaluate the delivery of an initiative*
    d. Outcomes – *the results and impacts of the initiative*

10. Evaluation can be focused on the impact of an initiative on outcomes (**summative**) or used to learn from the successes or failures of an initiative in order to shape its future direction (**formative**). In practice an evaluation is usually both.

11. The evaluation should identify and measure the direct and indirect benefits and costs of intervention over and above what would have happened in the absence of the intervention, i.e. establishing its **additional impacts**. It is important to define exactly what needs to be compared with what, i.e. where possible a 'control group' should be identified, to ensure differences in outcomes are not attributed to the intervention if they are in fact the result of changes in the 'state of the world'.

How is it to be evaluated?

12. The following should be considered when deciding how to evaluate an initiative; this section will deal with each in turn.
    a. Size of evaluation
    b. Data needed
    c. Methodology

Size of the Evaluation

13. Larger scale initiatives should have a bigger investment in the evaluation. Some basic monitoring of all initiatives should be undertaken as a matter of course, cost in relation to the overall spend on an initiative should be considered.

14. There is no official guidance on the proportion of an initiative's overall budget that should be spent on evaluation. This will need to be decided on a case-by-case basis, taking into consideration factors such as the duration of the initiative, the scale of the initiative, and the number of people affected. However, resources allocated to evaluation should be significant, perhaps 10% of the total budget and effort.

15. For larger initiatives it might be sensible to bring in external contractors who have experience in evaluating initiatives, however for smaller initiatives internal evaluation may be more cost effective.

Data Needed

16. It is important to consider how inputs and outputs etc. may be measured once identified. In using and collecting data, it is vital that the most relevant data are used and that these are collected in an appropriate manner. Data may be available through some of the following means:
    a. Monitoring/management information collected as part of the intervention.
    Monitoring information may yield important information about cost, number of participants and other administrative aspects of an initiative; however, this alone will be insufficient to fully evaluate the impact of an initiative on the target groups.

b. Existing sources – e.g. National Pupil Database (an annual census of all pupils in England).
Using existing sources of data may help to establish an impact and cut down on collection costs. It is important to investigate what kind of existing data may be useful; this will avoid duplicating data with additional collection and may aid comparability with other evaluations. However, it must be realised that for many initiatives, it is impossible to disentangle their effects and for some, the variables collected in existing datasets may be inappropriate – requiring bespoke data collection.

c. Additional data/evidence collected specifically for the evaluation e.g. interviews, focus groups, surveys and case studies.
Often existing data will not be available and other data will need to be collected by the evaluators, this may be a simple questionnaire after an event to try and determine the success of the initiative, however for larger scale initiatives interviews and surveys with key stakeholders will be important to an evaluation.

17. Collecting information from participants also provides them with an opportunity to feedback their views and experiences, giving policy makers the chance to record this feedback formally and to reflect it in how they develop the policy in future.

18. It is important to consider both qualitative and quantitative measures of the success of an initiative, and this should be reflected in the sort of information collected. Qualitative measures are those which provide descriptive assessment, with quantitative measures those which can be expressed as a numerical amount.

19. Longitudinal studies are surveys that get data form the same individuals at a number of points in time. They can provide a very good insight into the impact of an initiative as they are able to identify before and after outcomes for participant and non-participants in an initiative. They can be very costly and resource intensive to set up and run. There are, however, some longitudinal data sets which are available which may be able to do the same job. The National Pupil Database (NPD) is one example of a census level database of pupil characteristics and results from the age of 6 to 16.[2] Other cohort studies may also be used to investigate impacts of an initiative; the Youth Cohort Study is one example.

Methodology

20. Methodology is a key consideration; different initiatives will require a different approach and will be specific to the initiative in question, although most evaluations should include the following considerations:

a. Assessment (quantified if possible) of what happened.
It is important that there is a detailed understanding of the process of the initiative and how it worked or did not work. In some cases

[2] An NPD user group provides support for the access and use of NPD data: http://www.bris.ac.uk/Depts/CMPO/PLUG/index.htm

it might be possible to identify areas which worked particularly well for certain groups of people but not for others, explanations of this sort of outcome are necessary for a full and complete evaluation.

b. Comparison with target outturn.
The success of an initiative may be measured by comparing the actual output with a target, for example if the initiative aimed to get 60% of pupils in a cohort to a C grade at GCSE, how many actually achieved this target. The evaluation should also attempt to explain the reasons that this happened – this is important in terms of learning for the future. The measurement may also be attitudinal.

c. Comparative assessment of one or more counterfactuals – ideally using a control group.[3]
The counterfactual is the state that would have been reached in absence of some initiative. A comparison group can be valuable, if not always feasible, in allowing a full assessment of the success of an initiative, without a counterfactual being established it is not easy to separate the effects of the initiative with what may have happened anyway.

d. Assess the success of the initiative in achieving objectives, how initiative has (or has not) contributed to wider outcomes.
If outcomes are well identified at the beginning of the evaluation process then it should be possible to use data collected for the evaluation to identify where these outcomes have been achieved and also if there were further unexpected outcomes of the initiative. These should all be assessed to see if the initiative was successful. To facilitate this, benchmarking the situation at the start of the initiative may be useful. It is important to set both short-term and long-term objectives to give a more continuous view of how the intervention is developing and delivering outcomes. Focusing narrowly on one time horizon, whether immediate or in the future, risks creating a distorted view of effectiveness.

21. Another useful tool for thinking about the methodology of an evaluation is the Kirkpatrick model.[4] This model is helpful for thinking about how much evaluation to undertake for a particular initiative. There are four levels of potential impact of a initiative according to the Kirkpatrick model:

    a. Reaction – *The initial response to participation (e.g. immediate feedback on the initiative including things like enjoyment, facilities, best and worst aspects of initiative)*

    b. Learning – *Changes in people's understanding, or raising their awareness of an issue (this might include a comparison group to measure how things have changed as a result of the initiative, or use a baseline to establish changes)*

[3] For instance, comparing a group of participating pupils with an otherwise similar set of non-participating pupils.
[4] This is further detailed in the RCUK 'Evaluation: Practical Guidelines' document mentioned above.

c. Behaviour – *Whether people substantially modify what they do (longer term assessment of changes and measurement of the extent)*

d. Results – *To track long-term impacts of the initiative on measurable outcomes e.g. exam results (longer term more complex analysis – might be difficult to separate effects of an initiative with other things which have an impact on the relevant results)*

Feedback and Publication

22. The final report should summarise:

    a. How effective the initiative was in achieving objectives, why?

    b. Was initiative more beneficial for some groups than others, why?

    c. How could the imitative be improved to make it more effective?

    d. Cost effectiveness of initiative

    e. Why (if at all) outturn differed from that predicted in appraisal

    f. What results imply for future management or policy decisions.

23. Results and recommendations should feed into future decision making. It is at this point that resource devoted to managing and understanding the results of the evaluation really pays dividends. Efforts should be made to disseminate results widely; an important role of evaluation is to improve the efficiency and efficacy of future initiatives, both within the organisation which carried out the initiative but also in the wider community.

Examples of STEM Evaluations

The examples below set out a range of examples of the evaluation of STEM initiatives. They vary in scale and in type, including in and out of school experiences for pupils, and networking for CPD for teachers. Within the examples there is some brief commentary on what more could have been done to meet the guidelines.

Lab in a lorry

Background

Lab in a lorry offers 11-14 year-old pupils the chance to experience hands-on, interactive experiments in one of three laboratories which travel around the country. The scheme is funded by the Institute of Physics and the Schlumberger Foundation, and other regional partners. It aims to make physics and scientific careers more attractive to young people.

Evaluation Approach

The evaluation was conducted by the University of Durham. The evaluators issued questionnaires to pupils pre and post visit, including a group that would not visit the lorry to act as a comparison group, approximately 500 pupils responded. The questionnaires contained demographic questions and a series of statements to which pupils could indicate how strongly they agreed or disagreed using a five point scale.

Two weeks after the visit the evaluators interviewed some pupils using a semi-structure approach to find about their visit, their attitudes to science and career intentions.

The evaluation also included some observation of activities in the lorry and interviews with teachers to provide an additional perspective on the experience for young people.

Comments

Details of the budget for this programme were not available but it would seem that this is a cost-effective method of evaluation given the scale of the activity. The comparison group approach offers the chance to understand what difference the programme has made to the young people involved, and can record some of the information on outcomes that will be of interest to other partners, such as attitudes to science and career intentions.

Extending this example hypothetically, if trips to one of the lorries took place on a regular basis over a period of time, say a year or more, it might be worth recording which pupils attend regularly and comparing them to similar pupils using the National Pupil Database, to see if there is any boost to their exam results.

http://www.labinalorry.co.uk/about_lab_in_a_lorry/evaluation.cfm

After School Science and Engineering Clubs

Background

DCSF has provided £5 million of funding for 250 After School Science and Engineering Clubs to provide additional resources for pupils who show promise in these subjects. The clubs are aimed at:

- enriching, enhancing and extending the key stage 3 curriculum
- improving attainment in, interactions with and experiences of science among those pupils already showing interest and ability in these subjects
- encouraging these individuals to consider continuing their education in STEM
- improving collaboration between schools, and between schools and industry and the research base

Evaluation Approach

DCSF commissioned Sheffield Hallam University in November 2007 to undertake an evaluation of the After School Science and Engineering Clubs, to last one year. The evaluation is employing a before and after survey of 1000 pupils. Non-club members in participating schools are being used as a comparison group, although they are not a perfect match. The questionnaires record information on pupil characteristics and their attitudes to science and their future intentions.

The evaluation is also interviewing teachers and other staff in a representative sample of participating schools and non-participating schools to look at delivery (where relevant) and outcomes, such as pupil attitudes. There are other elements to the evaluation, including stakeholder interviews and case study visits to schools identified as having varied and interesting ways of working.

DCSF has, through the pupil survey mentioned above, collected the names of pupils so that their progress and attainment and subject choice in future years can be tracked in order to measure any impact. (This was done with the full knowledge of the pupil's parents by using an opt-out letter). As exact measures of future impacts on attainment and subject choice will necessarily not be clear for some time, it is intended that this will be carried out using our in-house analysts as and when attainment data becomes available.

Comments

This evaluation is in keeping with the scale of the initiative. There are a couple of interesting points to note. The within school comparison group is not always ideal, but it is easier, less costly and often more successful than trying to recruit a different set of schools to help. Other possibilities include using schools in future waves of a policy which have not yet received funding and implemented the policy.

The second interesting point to note is planning ahead to record the details of participating pupils to link the National Pupil Database. Again this is much cheaper than doing it retrospectively.

The evaluation is not explicitly looking at cost effectiveness, but could have addressed this issue using management information, perhaps focusing on the most significant areas of expenditure. Providing this information could have been made a condition of funding, although some resource would need to be devoted to checking the validity and consistency of data returned.

http://www.dfes.gov.uk/research/programmeofresearch/projectinformation.cfm?projectId=15264&keyword=engineering&keywordlist1=0&keywordlist2=0&keywordlist3=0&andor=or&type=0&resultspage=1

National Centre for Excellence in the Teaching of Mathematics

Background
DCSF is putting £15 million of funding the National Centre for Excellence in the teaching of Mathematics, which aims to raise the demand for CPD amongst maths teachers; to improve their subject knowledge; and to help coordinate a supply of good quality CPD to meet the demand. It consists of a web portal where teachers can access information and support and a system of regional network coordinators. A one year evaluation is being conducted by GHK and partners.

Evaluation Approach
The evaluation consists of interviews with staff members from the national and regional centres, which consist of an initial and follow up visit, to discuss delivery of the programme and how it has met its objectives.

There is also a review of the web portal, which is looking at who is using it and how are they using it. This is accompanied by a survey of web portal users, which is looking at similar themes, plus how useful it is.

There are also interviews with stakeholders including providers and schools and colleges, to find out the extent to which they are aware of and make use of the centre.

Comments
At the heart of the policy is networking, with the goal of helping teachers take up CPD opportunities. The impact of networks typically proves to be a very difficult thing to evaluate. For a policy larger in scale it may have been useful to expand the scope of the survey to understand how the centre is reaching those teachers who may not have undertaken much CPD recently, both through the portal and the networks.

It is likely to be very difficult or impossible to establish quantitatively what difference use of the centre has made to pupils, and hence to report impacts in terms of attainment and progression. It may be enough to understand what is different about practice, for instance through self-evaluation, and to seek feedback on pupil engagement and enjoyment of mathematics.

It is possible to conduct before and after analysis of impact but this is likely to be very costly and is unlikely to be realistic about findings that might be expected.

http://www.dfes.gov.uk/research/programmeofresearch/projectinformation.cfm?projectId=15222&type=3&resultspage=81

STEMNET (previously SETNET)

Background

This is a UK-wide Network aims to ensure that more young people in the UK make a choice to enter STEM related careers, and future generations are properly informed about science and technology. It aims to achieve this:

- By bringing STEM activities, experiences and excitement into classrooms throughout the UK, enhancing and enriching the national STEM curriculum.
- By linking those companies and other organisations that employ STEM educated people, and schools, in such a way that young people can get a clear idea of the diverse and exciting range of careers available to them.

Current funding (April 2005 - March 2008) from the Department for Innovation, Universities and Skills (DIUS) is £12.7 million, contributing to core operations at national, regional and local level.

Evaluation Approach

Commissioned from the Tavistock Institute, the evaluation was a pilot use of their Evaluation Framework for Science in Society initiatives, published in March 2007. [1]
The approach included:

- An assessment of STEMNET's objectives and activities, within the broad context of Science and Society policy and practice.
- A review of organisational structure and practices.
- An in-depth case study in London based on a SETPOINT and its interface with the local community.
- A survey of STEM teachers with varied level of STEMNET contact, about their awareness, interactions with and experience of STEM initiatives.
- Surveys of school students
- A summative assessment of the contribution STEMNET is making to policy and practice, and the outcomes and potential impacts.

The evaluation used triangulation of different types of evidence from different sources, to arrive at valid evaluation findings and conclusions. It was a mixed-method approach (methodological triangulation) drawing on a number of different data sources (data triangulation) and carried out by an experienced evaluation team (researcher triangulation).

A quantitative, quasi-experimental method and a qualitative, case-based method collected different types of evidence from a wide range of sources: primary data (e.g. from stakeholder interviews, observation of project participants, questionnaires and measures of scientific literacy); and secondary data (e.g. the content of evaluation reports and other material produced by the projects, and existing statistical data). These different views of STEMNET enabled the evaluation team to build a coherent picture.

Comments

There are issues around demonstrating causal relationships between STEMNET's activities and the perceptions and decision making behaviours of teachers and students, as there are many 'intervening factors' that may have an effect on behaviour.

STEMNET is a complex national network, and the STEM initiatives have different purposes – awareness raising; curriculum development; skills development; brokerage. With more time and resources, evaluators could have obtained feedback from a wider range of schools and stakeholders, to give a richer picture of the achievements and issues arising.

This pilot enabled the Evaluation Framework for Science in Society to be tested, and areas were identified for update, e.g. longitudinal studies, modelling of the initiatives, and guidelines on overcoming practical evaluation problems.

www.stemnet.org.uk
http://www.tavinstitute.org/work/journals/reports.php

The UK Resource Centre for Women in Science, Engineering and Technology

Background

The UK Resource Centre (UKRC) is funded by the Department for Innovation, Universities and Skills (DIUS), formerly the Office of Science and Innovation, with £6.9m from 2004 to 2008. Its aim is to increase the numbers of women pursuing professional careers in STEM, and to increase the numbers of women returning to those careers, and reaching senior positions in STEM-based areas of employment. UKRC was set up to provide advice to education professionals in Further and Higher Education, to employers and to women seeking STEM careers.

Evaluation Approach

The evaluation was commissioned by DIUS in December 2007 from the Tavistock Institute, who are applying the Evaluation Framework they produced for the Science and Society initiatives. The evaluation reflects the different positions, perspectives and backgrounds of a wide spectrum of stakeholders (staff, partners, sponsors, beneficiaries). The focus is:

- whether UKRC's mission, purposes and objectives – and the delivery models it adopts - are consistent with and support policy objectives aimed at attracting and retaining female scientists and engineers in STEM subjects and careers;
- whether a centre is the most appropriate and effective vehicle for delivering these policy objectives;
- to assess the operational performance of UKRC, and how this might be improved;
- to assist UKRC, DIUS, and all the other actors to learn from the experiences of UKRC, and so contribute to the development and implementation of policy.

The approach incorporates a mix of evaluation tools and methods, both quantitative and qualitative:

- Desk research – collecting and analysing existing data mainly compiled by the Centre, including statistics on uptake of activities; occupational pathways of Centre users; financial data; content analysis of UKRC outputs;
- Surveys – involving self-administered questionnaires targeting participants and beneficiaries, delivered by e-mail and on-line;
- Interviews and Focus Groups – with programme managers, beneficiaries, partners and other stakeholders;
- Case studies – both individual and thematic;
- Developmental activities – reviewing and reflecting on the ongoing evaluation results, and involving stakeholders in learning from them.

Comments

The evaluation approach outlined is comprehensive, and covers many aspects of the internal operation of the policy, which is sensible for a scheme of this size. It does seem to lack objective, quantifiable assessment of the impact of the policy on the numbers of women in STEM occupations– something which may have been difficult to capture.

The evaluation is due to complete in May 08.
Weblink to info about the scheme http://www.ukrc4setwomen.org.uk