



An Evaluation Framework for Health Promotion: Theory, Quality and Effectiveness¹

ERICA WIMBUSH

Health Education Board for Scotland

JONATHAN WATSON

Health Education Board for Scotland

There is increasing demand for evaluation work funded by public agencies to become more focused on demonstrating effectiveness. Focusing evaluation on outcomes and effectiveness meets the information needs of strategic planners and policy makers, but other stakeholders involved in managing, delivering or using public services and programmes may use other assessment criteria, such as improving the quality of programmes or programme design. The necessity and value of these other criteria are in danger of being obscured. Acknowledging the legitimacy of the range of stakeholder perspectives, this article presents a framework for evaluation that has been developed over a number of years within the context of evaluating health promotion programmes as part of the work of a national health promotion agency. It argues for an approach to evaluation which recognizes the contributions of theory and quality as well as effectiveness in programme development. The Health Education Board for Scotland (HEBS) framework for evaluation – and the analysis that informed it – demonstrates that there are many stages and forms of evaluation which contribute to the development of effective interventions. While outcome evaluations and effectiveness reviews tend to be the prized evaluation products for those concerned with policy and strategic planning, these forms of evaluation are just ‘the tip of the iceberg’ of what is required to build a sound evidence base, bringing together the full range of evaluation needs from the perspectives of all the different stakeholder groups.

KEYWORDS: effectiveness; evaluation; health; health promotion; quality; theory

Introduction

There is increasing demand for evaluation work funded by public agencies to become more focused on demonstrating effectiveness. This stems from both the move to make performance measurement within the public sector more outcome-oriented as well as the move to make policy making and practice more rational and 'evidence-based'.

The Cabinet Office White Paper *Modernizing Government* (Cabinet Office, 1999) clearly sets out the new drive towards improving the quality, efficiency and effectiveness of public services. In part, this entails an intensification of performance monitoring and evaluation, 'shifting the focus decisively from inputs to the outcomes that matter' (chapter 4, para. 6). A shift in emphasis towards demonstrating effectiveness and outcome-oriented evaluations is echoed in the new performance measurement and reporting procedures outlined for publicly funded agencies (National Audit Office, 2000). The new NAO report suggests that good practice in performance reporting involves a more comprehensive view of performance, including reporting the outcomes of activities and the information needs of stakeholders.

This emphasis on outcome-oriented evaluation and producing evidence of effectiveness is also apparent in the move to ensure that practice and decision making are 'evidence-based'. For example, within the health sector, enthusiasm for evidence-based medicine has spilled beyond the boundaries of clinical practice to incorporate health interventions in the community, which are concerned with improving population health and reducing inequalities in health. The aim of evidence-based medicine (and associated initiatives such as the Cochrane Collaboration) is to improve 'the conscientious, explicit, and judicious use of current best evidence' in planning and decision making about patient care (Sackett et al., 1996). The systematic review process is used to provide reliable and rigorous evaluations of the effectiveness of different treatments. Criteria for the inclusion of studies usually use randomized controlled trials as the gold standard for judging whether a treatment is effective. These clinically oriented criteria and methods have been directly transferred to the development of evidence-based practice in health promotion. In seeking to tackle the root causes of ill health and reduce inequalities in health outcomes, health promotion interventions adopt a 'whole systems approach', where cross-sectoral partnerships and high levels of community and user involvement are essential characteristics. What should count as evidence of effectiveness when it comes to such complex health interventions is a highly contested issue and has provided fuel for ongoing debates between positivists and constructivists concerning appropriate evaluation methodologies (Black, 1996; Tones, 1996; Speller et al., 1997, 1998a; Oakley, 1998a, 1998b; Platt, 1998; Nutbeam, 1999a, 1999b). This debate has now extended to include discussions about the role of quality assurance in the development of effective health improvement strategies (Speller, 1998; Davies and Macdonald, 1998). In addition, within the wider evaluation community, new theory-based approaches to evaluation have added a further dimension to the debate, emphasizing the importance of understanding the processes and mechanisms of change within programmes as

well as the outcomes desired and achieved (Chen, 1990; Connell et al., 1995; Pawson and Tilley, 1997).

Focusing evaluation on outcomes and effectiveness meets the information needs of strategic planners and policy makers, an important and powerful stakeholder group. Those stakeholders involved in other parts of the implementation chain – managing, delivering or using public services and programmes – may use other assessment criteria, such as improving the quality of programmes or programme design. The necessity and value of these other criteria are in danger of being obscured.

Involving stakeholders in the evaluation process is already recognized as good practice in evaluation at the level of individual projects or programmes (Quinn Patton, 1982), and tools for designing evaluations to address the diverse agendas of different stakeholders have been developed (Beywl and Potter, 1998). However, inter-sectoral collaboration on evaluation is not yet a feature at the national level in the UK. In the current policy environment, the need for greater orchestration of evaluation work and approaches across sectors is particularly acute given: (a) the multiple and overlapping policy initiatives which all require several levels of evaluation (strategic/national, programme and project levels); and (b) the emphasis on partnership funded initiatives and inter-agency collaboration. To avoid the situation where local partnership funded projects need to produce multiple monitoring and evaluation reports for multiple funders, it should be possible to develop a single comprehensive evaluation plan to serve all funders. Uncoordinated monitoring, evaluation and consultation work becomes a burden not only for local projects, but also for those targeted socially excluded groups and communities that are already showing signs of research fatigue.

Acknowledging the legitimacy of the range of stakeholder perspectives on what is valued and needed from evaluation, in this article we present a framework for evaluation that has been developed over a number of years within the context of evaluating health promotion programmes as part of the work of a national health promotion agency, the Health Education Board for Scotland (HEBS, 1999). We argue for an approach to evaluation which recognizes the contributions of theory and quality as well as effectiveness in programme development. The HEBS framework for evaluation is presented in the final section of the article.

What Sorts of Evaluations are Needed and Valued?

According to Weiss (1999), the overall aim of evaluation is to assist people and organizations to improve their plans, policies and practices on behalf of citizens. While it is relatively easy to build consensus around evaluation for learning and improvement, there are important differences, in perspective and in emphasis, among stakeholder groups around what forms of evaluation are needed and valued. This can be illustrated with reference to the field of health promotion (IUHPE, 1999: 3; Watson and Platt, 2000; Wimbush, 1999).

Policy Makers and Strategic Planners

Policy makers and strategic planners need to be able to judge the effectiveness, or likely effectiveness, of health promotion programmes in order to make decisions about the most efficient and effective deployment of public resources, decisions for which they are accountable to elected representatives and citizens. The sorts of questions they need answered are 'what works?' or 'what are the best buys?' Systematic reviews of effectiveness are intended to provide answers to such questions but tend to draw only on evidence from experimental and quasi-experimental research designs. They also require economic evaluations of health promotion interventions which look at the relationship between inputs/investments and short-term health gains.

Programme Managers

Programme managers who are budget holders responsible for the delivery of health promotion programmes and local health strategies in 'real-life' circumstances need evaluations which provide feedback on the success of a range of different projects and initiatives and the extent to which they contribute to the achievement of local strategies. Here success is most likely to be assessed in terms of achieving defined objectives, reaching the targeted populations and the extent to which the local partnerships are sustainable.

Practitioners

Practitioners who are responsible for the operation and running of community health projects and services, often involving local partnership funding, find evaluations most useful when they engage with the practicalities of the implementation process, and provide feedback from people and other agencies involved in collaborative action. Evaluations which play a developmental or formative role, identifying areas for change or improvement, are particularly valued. However, those working in local projects often perceive funders' requirements for monitoring and evaluation as a 'top down' demand and struggle to cope with the multiple, duplicative and sometimes contradictory evaluation requirements of different funding bodies.

Community Groups/Users

The population likely to benefit from the service or programme (e.g. clients, users, the community) will be concerned with the quality of service provision, the extent to which it is relevant to their perceived needs, and the extent to which its operation is participatory or consultative. They are most likely to value evaluations which provide an avenue for feedback and involvement, address quality issues and assess community/user concerns and satisfaction. Whether an initiative delivers tangible benefits for the community is a form of effectiveness evaluation that is likely to be valued by local people, whether or not they form part of the target population.

Professional Evaluators

Professional evaluators (including academic researchers) tend to engage with evaluation as a knowledge-building exercise, seeking to improve knowledge and

understanding of the relationship between an intervention and its effects. They are also concerned to maintain quality standards for research, in particular with regard to research design, methodological rigour, reliability and validity. However, evaluators employed within health promotion practice settings are often frustrated by being expected to 'evaluate everything' on a small budget and not having the resources to conduct what they regard as 'quality' research. Academic researchers are often highly critical of the quality of evaluation research carried out in practice settings, but are sometimes all too ready themselves to conduct resource-intensive evaluations of effectiveness with little attention to assuring the quality of the intervention being tested. This situation contributes to findings from large-scale evaluations which demonstrate the failure of community health interventions (e.g. Stanford, Pawtucket, Minnesota, Heartbeat Wales), the failure being attributed to the quality of programme implementation and delivery.

Inevitably, there is likely to be some overlap between the interests of the different stakeholder groups. In advocating the need for evaluation evidence that is relevant to their own particular priorities, the different stakeholder groups can disregard the necessity and contributions of other forms of evaluation. This suggests a need for more 'joined-up' thinking and partnership working on evaluation across the different stakeholder groups – policy makers and strategic planners, programme managers and practitioners, user/consumer groups – as well as those commissioning and doing evaluation work.

Building a Common Framework for Evaluation

In addition to the principle of involving key stakeholders in the evaluation process, the above stakeholder analysis suggests that a common framework for evaluation should also address information needs around effectiveness, quality and the implementation process.

Accountability and Effectiveness

The link between evaluation and public accountability is as strong as ever, bringing an imperative for evaluations to address effectiveness so as to demonstrate both the intended and unintended consequences of policy initiatives and programmes. The requirement to focus on effectiveness has brought a spate of initiatives concerned with developing indicators. For example, in the US a community indicators movement has arisen as local government, businesses, and community leaders form partnerships for improving community life. Sets of community indicators which encompass a community's economic, environmental and social well-being are tracked over time, as a way of evaluating what progress and changes have occurred in the community in relation to such agendas as sustainability, healthy communities or quality of life (see http://www.rprogress.org/progsum/cip/cip_main.html). However, while the tracking of indicators of desired outcomes is necessary, they are not sufficient in themselves. Outcomes indicators need to be located within a model or framework which allows the linking of the actions planned to the desired outcomes over time (Ralls and Thomson, 2000).

Evaluation 6(3)

Having a clear model for outcomes helps to shape the expectations of stakeholders about what a programme can be reasonably expected to achieve over a defined time period.

In the field of health promotion, the modelling of outcomes from health promotion actions for evaluation purposes has been tackled mainly by two writers – Keith Tones in the UK (Macdonald et al., 1996; Tones, 1998) and Don Nutbeam in Australia (Nutbeam, 1996, 1998; IUHPE, 1999: 6).

For Tones, the relationship between the initial inputs and eventual outcomes of a health promotion initiative is considered in terms of a chain of complex and interacting interventions that occur over time, sometimes quite extensive periods of time. This is referred to as a ‘proximal-distal’ chain of effects. The greater the number of links in this proximal-distal chain, the greater the effort which will be needed to attain the ultimate goal and the less likely it is that success will be achieved. Given the complexities of many health promotion programmes and the distribution of effects over often quite lengthy periods of time, Tones argues that three types of indicators are needed:

- indirect indicators (Time 1) which indicate changes that are a direct result of a health promotion intervention;
- intermediate indicators (Time 2) which indicate changes that are intended to follow on from a health promotion intervention;
- outcome indicators (Time 3, Time 4, Time 5) which indicate more distant changes in health behaviours, service use and health status.

Don Nutbeam provides a framework for defining the outcomes associated with health promotion activity. Three broad areas of health promotion action are defined (education, social mobilization and advocacy) which are linked in a dynamic relationship to a hierarchy of outcomes: the immediate health promotion outcomes (programme impact measures); intermediate health outcomes (modifiable determinants of health); and the desired long-term health and social outcomes (reductions in morbidity, avoidable mortality and disability, improved quality of life, functional independence and equity).

Implicit in both these models is the notion that changes unfold over time and that outcomes need to be differentiated on the basis of a time dimension. These models also both emphasize that health promotion programmes often involve a diverse range of actions aimed at different levels: for example, educational forms of action that seek change at the level of the individual; community development projects whose efforts are concerned with community empowerment and improving the quality of community life; advocacy approaches that may seek changes in environments or legislative reform. Such diversity in the range of possible actions and in the outcomes sought can make the link between health promotion actions and eventual health outcomes complex and difficult to track. This is particularly the case where there are multiple level actions and where the time lapse between health promotion actions and outcomes is extensive.

However, neither of the above models takes into account the ‘capacity-building’ role of health promotion and its related outcomes. Hawe et al. (1997) argue that in addition to assessing and measuring the health gain related outcomes of

health promotion programmes, we need to assess the value and outcomes of the less visible capacity building process.

Health promotion programs should be thought of as an investment, the benefits of which are not represented in full by the health outcomes delivered from programs immediately on the completion of the funding period. An indication of how good an investment might be in the longer term comes from the capacity building indicators. To assess the value of a health promotion program in terms only of the 'amount' of health gain seemingly delivered would be like using a ruler to measure a sphere. Capacity-building alongside, or as a prelude to, program development, implementation, evaluation and maintenance represents a 'value-added' dimension to health outcomes. (Hawe et al., 1997: 38)

This raises one of the core tensions for evaluation between assessing effectiveness and demonstrating the achievement of results (e.g. improving health-related outcomes), and understanding the mechanisms and processes whereby such results were achieved, including for example the additionality that might stem from the greater capacity of a community or organization to take action to address health issues in the future.

In the field of health promotion research, this tension is manifest in a long-established debate between those who subscribe to a classical experimental paradigm and seek to measure the size of outcomes and those who seek to explain outcome patterns by accommodating qualitative and contextual data in their evaluations. These debates show signs of moving beyond the simple dichotomies drawn between qualitative and quantitative methodologies and between the virtues of process versus outcomes evaluation towards a more sophisticated, pluralistic approach to the evaluation of health promotion interventions (IUHPE, 1999; Davies and Macdonald, 1998; Scott and Weston, 1998; Watson and Platt, 2000). Examples of randomized control trials of health promotion interventions which incorporate a detailed exploration of process are: the evaluation of SHARE, a teacher-led sex education programme in secondary schools in Scotland (Wight, 1997); and the evaluation of PRISM, an integrated primary care and community development intervention for mothers of new babies in Australia (Lumley, 1997).

Programme Theory – Reconciling Processes and Outcomes

Understanding the process of implementation and the mechanisms by which certain outcomes will be achieved is the point at which evaluation enters the domain of programme 'theory' (Chen, 1990). Evaluators often start out by clarifying a programme's aims, objectives and the desired outcomes, but theory-based approaches suggest that evaluators also go on to elicit the key assumptions and linkages underlying how a programme has been designed, i.e. understanding the 'logic' of how the programme is supposed to operate to achieve the desired outcomes. There are two main theory-based evaluation approaches which have become influential – 'theories of change' (Connell et al., 1995) and 'realistic evaluation' (Pawson and Tilley, 1997).

The US Round Table on Comprehensive Community Initiatives for Children and Young People developed a theory-based approach to the evaluation of

Evaluation 6(3)

comprehensive community initiatives which is referred to as the Theories of Change approach (Connell et al., 1995). Weiss (1995) suggests that all programmes have explicit or implicit 'theories of change' about how and why a programme will work, that influence decisions around programme design. Once these theories of change have been made explicit, they can drive the development of an evaluation plan that tests whether the programme's theory holds up when the programme is implemented. Connell and Kubisch (1996) describe the Theory of Change approach as the articulation and testing of a programme's desired outcomes, and the timescale for these to be achieved, together with the processes whereby these will be arrived at, making adjustments to methods and goals along the way. It is thus an approach to evaluation which is seen as reconciling processes and outcomes (Hughes and Traynor, 2000). In the UK it is an approach that has been applied, with encouraging early results, to partnership-based community initiatives that are seeking to improve health (among other things). Examples include the evaluation of Health Action Zones in England (Bauld and Judge, 1998; Judge et al., 1999) and the evaluation of an Anti-Poverty Strategy implemented in eight localities using community development approaches (Hughes and Traynor, 2000).

Another form of theory-based evaluation is Realistic Evaluation (Pawson and Tilley, 1997). Programme theories are framed in terms of propositions about the 'Context + Mechanism = Outcomes' configuration of a programme, i.e. how Mechanisms are fired in certain Contexts to produce certain Outcomes. The theories underlying a programme's design are generated through a detailed analysis of the programme in order to identify what it is about the measure which might produce change, which individuals, sub-groups and locations might benefit most readily and what social and cultural resources are necessary to sustain the changes. The evaluation then tests these hypotheses about the programme. This approach to theory-based evaluation has been widely taken up, including a series of evaluations of initiatives within health and social services carried out by the Centre for Evaluation Studies at the University of Huddersfield (Kazi and May, 1999).

Quality Assurance – Improving Chances of Effectiveness

A further tension that exists between the stakeholder interests of researchers and practitioners relates to the lack of attention given by researchers to the quality of the programme, service or organization which is to be evaluated. As part of the *Modernizing Government* agenda, the development of quality assurance in the public sector is being given high priority at present. In terms of evaluation, quality assurance is relevant in the sense that it supports the development of a systematic and reflective approach to improving practice and performance. It is also a critical component of process evaluation since in trying to understand and explain outcome patterns, it is important to be able to distinguish between the effects of the intervention itself and the quality of the delivery.

In health promotion, it has been argued that successful outcomes are unlikely to be delivered without attention to the quality of the intervention (Speller, 1998; Speller et al., 1998b), where quality covers the principles of participation and

partnership as well as the processes of programme planning, design and delivery. Speller suggests that quality assurance is a process that is not separate from, but complementary to, the evaluation of health promotion programmes:

Quality assurance and effectiveness research in health promotion are not separate endeavours, but are interwoven in an ideal process of intervention testing and delivery. Health promotion practitioners have a central part to play in the design of interventions for testing, in particular to ensure that issues relating to consumer perceptions of quality have been considered which will maximise acceptability and professional views are built in, primarily to ensure that implementation is feasible. (Speller, 1998: 88)

Quality assurance requires having systems in place to define desirable and achievable standards of practice or performance, monitoring and regularly reviewing current practice/performance to check if standards are being reached, and taking action to enable standards to be achieved (Evans et al., 1994). This cyclic process of monitoring and reviewing agreed standards is thought to optimize the likelihood of efficiency and effectiveness.

Stages of Programme Development

Implicit in the above is that evaluation (in which monitoring, review and other quality assurance processes are included) contributes to the development and design of programmes. The relationship between programme and evaluation can be further elaborated by thinking about the role and focus of evaluation at different stages in the development of a programme or project. A lifecycle framework and a cybernetic or feedback model are two examples of frameworks used to help people think about the relationship between programme and evaluation (EDRU, 1992).

In health promotion, a number of planning models have been devised to assist practitioners in the planning, development and delivery of health promotion programmes:

- Australia – The Health Workers Guide (Hawe et al., 1990);
- The US – The Five Stage Community Organization Model (Bracht and Kingsbury, 1990);
- Canada – The PRECEDE-PROCEED Planning Framework For Health Promotion (Green and Kreuter, 1991);
- Australia – The Staged Approach To Health Promotion (Sanson-Fisher and Campbell, 1994);
- The Netherlands – The ABC Planning Model (de Vries, 1998).

All the above models identify a series of phases or stages, which are sometimes further sub-divided into steps. While each have their distinctive features, they also share a common understanding of the key stages involved (see Box 1).

Each of the above stages is paralleled by a different evaluation focus and set of evaluation questions (see Table 1). This illustrates how the issues of theory, quality and effectiveness can be appropriately addressed at different stages in the development and implementation of a programme. Assessing effectiveness too early in the life of a project will be wasted effort since outcomes are unlikely to

Evaluation 6(3)

Box 1

Key stages

1. Planning – an initial preparatory phase characterized by the needs assessment process.
2. Design – a phase when the aims, objectives, target group and actions are defined and the programme undergoes initial testing or piloting.
3. Implementation – a phase when the programme becomes fully implemented and outputs are delivered.
4. Continuation/diffusion/dissemination – a phase which includes time for review and reflection on learning from the evaluation process so that implications for future practice can be identified and actioned. It is also a stage where those programmes which have proved effective and warrant continuation or wider implementation are disseminated (or 'demonstrated') to ensure continuation and wider utilization.

be realized until a project is fully operational. Equally if outcomes are assessed towards the end of a project without appropriate quality assurance or process evaluation, the results are likely to be unhelpful in guiding future action because what generated the observed outcomes will remain unknown. The theory underlying how a programme was intended to work at a pilot stage can be tested during implementation and adjusted. If a successful programme is transferred to another population and setting and replicates similar outcomes, this will strengthen the generalizability of its theory of change.

HEBS Evaluation Framework for Health Promotion

The evaluation framework developed by HEBS (HEBS, 1999) uses the key stages of programme development as the basis for differentiating between the types of evaluation used and useful in health promotion practice. The HEBS framework identifies the different purposes of evaluation and the associated evaluation questions that are characteristic of each of these stages, acknowledging the importance of assessing effectiveness, as well as assuring quality and making explicit the mechanisms of change implicit in a programme's theory. The different types of evaluation identified in the HEBS framework are outlined briefly below and the full framework is given in Table 2.

Planning Stage: Systematic Reviews of Effectiveness

In the planning stage, once a health-related problem and the population group at risk have been identified, a second phase in the needs assessment process involves an option appraisal process which takes into account:

- (a) learning from other evaluation research about the most effective ways of addressing the problem with a particular group and/or within a particular setting (systematic reviews of effectiveness);
- (b) how the health-related need/problem is currently addressed by current policies and service provision (review of current provision/policy);

Table 1

Stage	Evaluation focus	Evaluation questions
Planning	Learning from other evaluations of effectiveness; option appraisal.	What are likely to be the best/most effective ways of addressing a particular need or problem with a particular group/setting?
Design and pilot	Feasibility of proposed approach; 'theory of change'.	Is the proposed programme feasible and acceptable? What outcomes can be realistically achieved in what time period? How and why will/does it work? How should the programme be adapted to maximize effectiveness?
Implementation • <i>Early start-up</i>	Delivery and quality assurance; monitoring and review systems; baselines.	Are we on track? Are there any problems that need to be addressed? What action needs to be taken to improve practice or performance?
Implementation • <i>Establishment</i>	Implementation process; reach; programme impacts/results.	How is the project working? Is it being implemented as intended? To what extent is the target population being reached? To what extent are programme objectives/impacts being achieved? At what cost?
Implementation • <i>Fully operational</i>	Intermediate outcomes/ effectiveness.	To what extent were intermediate outcomes achieved? How were these achieved? In which groups/settings are the greatest benefits shown?
Dissemination	Replicability of outcomes; generalizability of theory.	Can the programme be transferred to another setting or population and achieve the same outcomes?

Table 2. HEBS Evaluation Framework

Stage of project development	Type of evaluation	Purpose of evaluation	Research questions	Application	Whose responsibility and what resources are required?
Planning	Systematic reviews of effectiveness.	Overview of evidence of effectiveness from outcome evaluations around a specific topic, setting.	What are effective ways of addressing a particular need or problem? What makes a difference?	All (where possible).	International collaborations, government departments and national agencies.
Design and pilot	Developmental evaluation.	To assess the feasibility, practicability and acceptability of the new project and its processes /mechanisms and to test the potential effectiveness of a new approach.	Is it feasible and practicable? How does it work? What outcomes can realistically be achieved? How should it be adapted/refined to maximize effectiveness?	Projects in design or pilot stage which are testing new or innovative approaches.	Funders/sponsors of local programmes.
Implementation – early start up	Monitoring and review (for evaluation and quality assurance).	To monitor and review progress in achieving agreed milestones and agreed quality standards in order to improve quality and efficiency.	What have we achieved so far? How could we improve?	All.	Project managers. Cost of training in project management and quality assurance which should include review and monitoring procedures.

Table 2. Continued

Implementation – establishment	Impact evaluation.	To assess the short-term effectiveness of a project in terms of its reach and immediate impacts. If data on costs are available, simple economic evaluation measures can be produced.	How is the project working? Is it being implemented as intended? To what extent is the target population being reached? To what extent are objectives/goals being achieved? What impacts have been achieved? At what cost?	All large projects (total budget £100k+).	Funders/sponsors of local programmes. Funding for evaluation should be proportional to project costs. Guidelines suggest approx. 10–15% of total project costs.
Implementation – fully operational	Outcome evaluation.	To assess the longer-term effectiveness of a project using intermediate outcome measures. If data on costs are available, more complex economic evaluation measures can be produced.	To what extent were intermediate outcomes achieved? How were these achieved? In which group/settings are the greatest benefits shown?	Projects where short-term effectiveness has been assessed positively.	Source of funding is often research council grants, trust funds, or government departments. Applications are led by academic researchers with evaluation expertise, but developed in collaboration with programme team.
Dissemination	Transfer evaluation.	To assess the replicability of a project's mechanisms/processes and outcomes.	Can the project be transferred to another setting or population and achieve the same outcomes?	All large projects that have proved effective in one setting/population.	Local agencies who want to apply approach locally. Government agencies wanting to disseminate effective practice.

- (c) what professional 'experts' regard as the best ways of addressing these needs/problems (consultation interviews or seminar).

The most recent review of effectiveness in health promotion (Peersman et al., 1999) found over 450 systematic reviews had been conducted, covering some areas more extensively than others. According to Morgan (1997), the two main themes to have emerged from these reviews to date are: first, the shortage of evaluation studies of sufficient quality to be eligible for inclusion; and second, their inability to draw positive conclusions which advance practice. However, the criteria used to select evaluation studies for inclusion in these 'effectiveness reviews' have been the focus of heated debate (HEA, 1997), in particular the transfer of the clinically defined standards of research quality to other, non-clinical areas (Speller et al., 1997). The expectation that there might be universal solutions to an identified problem (e.g. teenage pregnancies) has been challenged by the advocates of realistic evaluation who argue that the question posed should not be simply 'what works?', but 'what works, for whom, in what circumstances?' (Pawson and Tilley, 1997).

Design and Pilot Stage: Developmental Evaluation

The effectiveness of interventions is increased if an initial pilot stage is undertaken before the proposed programme is fully implemented. A programme plan can be designed which is based on the initial assessment of need and appraisal of what is likely to be the most effective or 'best' intervention, given the evidence and resources available, and what can be achieved within a particular setting and set of partner agencies. Against this backdrop, the design stage involves defining the long-term goal of the programme, setting programme objectives, defining the range of activities required to meet these objectives, identifying staffing and training requirements, setting up administration, publicity and monitoring procedures.

Developmental evaluation is an essential part of this design stage. Formative evaluation is likely to be most appropriate approach since the prime purpose of the evaluation is developmental and the process is iterative, providing continuing feedback from key stakeholders and the target group/project users in order to adjust, refine and optimize the programme's focus, design and ultimate effectiveness. If the programme is found at this stage to be unfeasible or impracticable without major revisions, then the project should be abandoned and a new approach devised.

The purpose of evaluation in this stage is:

- to assess the feasibility, practicability and acceptability of the proposed programme through piloting on a small scale;
- to identify what impacts and outcomes are realistic for the programme to achieve over a defined period of time;
- to develop an understanding of how the programme will operate over its funding period in order to achieve these outcomes;
- to develop and test any educational materials with the representatives of the target population;
- to review and adjust the programme's initial design (aims, objectives,

activities, resources, timetable, outcomes) following the pilot stage in order to maximize its potential effectiveness.

This is the stage of programme development when Health Impact Assessment (HIA) might be appropriate. HIA is a form of prospective outcome evaluation advocated in the recent White Papers on Health (Department of Health, 1999; Scottish Executive, 1999; Welsh Office, 1998), concerned with estimating the potential population health effects (positive and negative) of non-health care initiatives so as to inform their planning and design, maximizing the health benefits and mitigating health risks.

Implementation Stage (Early Start-Up): Monitoring and Review

For evaluation purposes, it is helpful to distinguish between different phases of implementation: early start-up, establishment and a fully operational phase. Overall, the implementation stage is characterized by the operation of the full programme across all sites in its revised post-pilot form. The main tasks here are project management, quality assurance and evaluation.

At the start of a project, the project manager is concerned with defining appropriate milestones for the project, review cycles and agreeing with key stakeholders appropriate performance indicators and quality standards for the project. Monitoring and review systems should be set up to continue throughout the duration of the project's life for both evaluation and quality assurance purposes. These systems include:

- monitoring systems for routinely recording data about inputs, outputs, project activities and any agreed quality standards;
- evaluation work should begin by looking at management issues around the delivery of the project and quality assurance. If the impacts and outcomes of the project are to be assessed over time, it may be appropriate to collect baseline information at this early stage.

Implementation Stage (Establishment): Impact Evaluation

This phase of implementation is when the project has become stable, project staff have gained experience and confidence and early problems have been addressed. At this stage, 'impact evaluation' is appropriate and the evaluation focus turns to examining the implementation process: the extent to which the project is working as planned; how far the project has reached the target population; and the immediate effects of the project (i.e. its impacts or results) on the target population and others. If monitoring data on costs is available, simple economic evaluation measures such as cost effectiveness and/or cost:benefit ratio might also be produced.

Implementation Stage (Fully Operational): Outcome Evaluation

Once the project is well established, the evaluation can focus on effectiveness – whether the end results, or intermediate outcomes, are being achieved and thus the extent to which the project has been effective in contributing to longer-term health and social policy goals. Outcome evaluation should be conducted when an

Evaluation 6(3)

impact evaluation has already demonstrated a programme's short-term effectiveness, ideally in several settings/populations, but long-term effectiveness is still unknown.

To allow long-term follow-up over time, this type of evaluation requires dedicated and substantial research resources and those with specialist evaluation expertise who can advise on appropriate research designs and methods, implement these and conduct the appropriate analysis. One of the biggest problems with this form of evaluation is providing evidence of a causal link between the project being evaluated and the outcome measures. Experimental and quasi-experimental research designs go some way towards addressing this problem, although these designs are regarded by many as a research design that is neither feasible nor desirable for community-based interventions.

If the outcome evaluation indicates that an intervention has no significant effect overall, the process evaluation should indicate whether it was due to programme implementation or the evaluation. If due to programme implementation, failure may be because the assumptions underlying the programme were problematic or the programme was not fully or evenly implemented. If the outcome evaluation indicates that the intervention was effective, then the dissemination phase should also address the transferability of the intervention to other populations and settings.

Dissemination Stage: Transfer Evaluation

The dissemination stage begins when there is information available for dissemination beyond the immediate audience of project staff, funders and stakeholders, about the 'results' of, or learning from, the impact and outcome evaluation research. Typically, this is when the initial project funding period comes to an end.

Programmes that have proven to be effective will only have significant impact if they are disseminated and taken up more widely. This is the purpose of 'demonstration projects'. The focus of evaluation at this stage is on the transferability of the programme and the replicability and sustainability of its outcomes when transferred to a wider range of settings and/or populations. The main difficulty with programme transfer is being able to identify (a) which elements of the programme were effective and need to be transferred; and (b) what the necessary pre-conditions are for the programme to be effective. These problems are minimized if evaluations carried out in the earlier stages of a programme's development have created an understanding of the mechanisms that are most effective and of the necessary pre-conditions. Potvin (1996) disputes the proposition that the purpose of evaluating programmes in their dissemination phase is to show that the conditions necessary to produce the expected outcomes are implemented. She argues that the evaluation agenda should be broad and open in all phases of project development and that outcome assessment is needed throughout the entire cycle of any project's life.

The issues of replication and sustainability are crucial for demonstration projects, since these can only usefully inform policy and practice if the desired outcomes are generalizable to other settings and populations. The sustainability of outcomes will depend on the capacity of the system to prolong health promotion

programmes and thus to multiply health effects as well as to develop new programmes (Hawe et al., 1997).

Implications

The HEBS evaluation framework serves to demonstrate that there are many stages and forms of evaluation which contribute to the development of effective interventions. While outcome evaluations and effectiveness reviews tend to be the prized evaluation products for those concerned with policy and strategic planning, these forms of evaluation are just 'the tip of the iceberg' of what is required to build a sound evidence base, bringing together the full range of evaluation needs from the perspectives of all the different stakeholder groups.

The evaluation framework also points to the value of the evaluation products generated in practice settings, which are concerned with the development of quality forms of action that are relevant and acceptable to local populations. The framework helps to make more visible the different types of evaluation research that are necessary and appropriate to develop within the practice setting, and indicates in what areas of evaluation it might be appropriate for practitioners to take the lead and develop skills. In this sense the evaluation framework contributes to improving the fit between research and practice, an endeavour which lies at the heart of 'best practice' in health promotion (Nutbeam, 1996).

The evaluation framework is premised upon a systematic approach to programme evaluation, understanding not only the outcomes and effectiveness of health promotion programmes, but also *how* and *why* certain outcome patterns emerge. It is necessary for evaluation to contribute to developing an understanding of what Carol Weiss (1995) terms a programme's 'theory of change' and what Pawson and Tilley (1997) refer to as the context and mechanisms of an intervention. This articulation and testing of programme 'theory' are essential if successful interventions are to be transferred to other settings.

In attempting to sketch out the big picture (i.e. all the evaluation pieces needed to build up a sound knowledge base for health promotion), the evaluation framework helps to clarify the division of evaluation labour, distinguishing between the different forms of evaluation required at different stages in the programme development process. An important distinction is between:

- project level self-evaluations led by the project themselves and resourced by project funding; this would include developmental evaluation, monitoring and review for quality assurance purposes and impact evaluations; and
- the rather more complex evaluations that academic researchers or specialist evaluation consultants are commissioned or funded to undertake with national funding; this would include outcome evaluations, transfer evaluations and systematic reviews of outcome evaluations.

This distinction is important for two reasons. First, it helps to create a more realistic expectation among funding bodies of what monitoring and evaluation are appropriate to conduct at project level. Second, for professional development purposes, it serves to highlight what evaluation capabilities are needed nationally and locally to fulfil these evaluation requirements.

Conclusions

In this article we have attempted to draw together an analysis of the different stakeholder perspectives on the needs and value of evaluation in a way that highlights the importance of assessing effectiveness, while also assuring quality and making explicit the mechanisms of change implicit in a programme's theory. Each of these foci for evaluation are appropriate at different stages in a programme's development. This forms the rationale underlying the HEBS evaluation framework for interventions seeking to improve population health, an endeavour which inevitably involves a whole systems approach and inter-agency partnership working. The framework is based on the assumption that it is desirable and possible to adopt a systematic approach to the planning, design, implementation and dissemination of programmes. Indeed, this is also the cornerstone of what is termed 'evidence based health promotion' (Wiggers and Sanson-Fisher, 1998). To achieve this requires a more 'joined-up' approach to programme development and evaluation across the domains of policy making, planning, practice and research. For example, this framework suggests that investments in long-term and intermediate outcome evaluations should not be undertaken unless there is evidence that the intervention is based on a tested theory of change and has proven short-term effectiveness. At present there is no research-practice linking mechanism through which successful programmes that have been tried and tested in a community setting can be nominated for further outcome evaluation research.

A more orchestrated approach to evaluation across sectors and funding agencies might also help to address two further problems: first, how to maximize the learning potential from the UK's wealth of monitoring and evaluation data coming back from the many new inter-sectoral policy initiatives, such as Health Action Zones, Social Inclusion Partnerships, Healthy Living Centres, New Deal; and second, how best to build national and local level capacity for the sort of performance measurement, monitoring and evaluation work demanded within these programmes.

Acknowledgements

This framework is built upon learning from the experience of evaluation at HEBS over the last six years, and discussion with colleagues in HEBS and local health boards, as well as those involved in monitoring and evaluation work in other agencies and sectors. In particular, we would like to acknowledge the contributions of colleagues in the Research and Evaluation Division, and the valuable feedback from Nick Tilley and Sandra Nutley and other anonymous reviewers on early drafts of the article.

Note

1. The opinions expressed in this publication are those of the authors and not necessarily those of the Health Education Board for Scotland.

References

- Bauld, L. and K. Judge (1998) 'Health Action Zones', *Health Variations Newsletter* 2: 10-1.
- Beywl, W. and P. Potter (1998) 'RENOMO - A Design Tool for Evaluations: Designing Evaluations Responsive to Stakeholders' Interests by Working with Nominal Groups using the Moderation Method', *Evaluation* 4(1): 53-72.
- Black, N. (1996) 'Why We Need Observational Studies to Evaluate the Effectiveness of Health Care', *BMJ* 312: 1215-8.
- Bracht, N. and L. Kingsbury (1990) 'Community Organization Principles in Health Promotion: A Five Stage Model', in N. Bracht (ed.) *Health Promotion at the Community Level*. Newbury Park, CA: Sage Publications.
- Cabinet Office (1999) *Modernizing Government*, Cmd 4310. London: The Stationery Office.
- Chen, H. T. (1990) *Theory-Driven Evaluations*. Thousand Oaks, CA: Sage Publications.
- Connell, J. P. and A. C. Kubisch (1996) *Applying a Theories of Change Approach to the Evaluation of Comprehensive Community Initiatives*. Washington, DC: The Aspen Institute.
- Connell, J. P., A. C. Kubisch, L. B. Schorr and C. H. Weiss (eds) (1995) *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*. Washington, DC: The Aspen Institute.
- Davies, J. K. and G. Macdonald (eds) (1998) *Quality, Evidence and Effectiveness in Health Promotion: Striving for Certainties*. London and New York: Routledge.
- Department of Health (1999) *Saving Lives: Our Healthier Nation*. London: The Stationery Office.
- De Vries, H. (1998) 'Planning and Evaluating Health Promotion', in D. Scott and R. Weston (eds) *Evaluating Health Promotion*. Cheltenham: Stanley Thornes Ltd.
- Evaluation, Development and Review Unit, Tavistock Institute (1992) *A Guide to Local Evaluation. A Resource Produced for the Enterprise in Higher Education Initiative*. London: Employment Department.
- Evans, D., D. Head and V. Speller (1994) *Assuring Quality in Health Promotion: How to Develop Standards of Good Practice*. London: Health Education Authority.
- Green, L. and M. Kreuter (1991) *Health Promotion Planning: An Educational and Environmental Approach*. Toronto and London: Mayfield Publishing Company.
- Hawe, P., D. Degeling and J. Hall (1990) *Evaluating Health Promotion: A Health Workers Guide*. Sydney, Philadelphia, PA and London: MacLennan and Petty.
- Hawe, P., M. Noort, L. King and C. Jordens (1997) 'Multiplying Health Gains: The Critical Role of Capacity Building Within Health Promotion Programs', *Health Policy* 39: 29-42.
- Health Education Authority (1997) *Reviews of Effectiveness Workshop: Their Contribution to Evidence-Based Practice and Purchasing in Health Promotion*. London: RSA.
- Health Education Board for Scotland (1999) *Research for a Healthier Scotland: The Research Strategy of the Health Education Board for Scotland*. Edinburgh: HEBS.
- Hughes, M. and T. Traynor (2000) 'Reconciling Process and Outcome in Evaluating Community Initiatives', *Evaluation* 6(1): 37-49.
- International Union for Health Promotion and Education (1999) *The Evidence of Health Promotion Effectiveness: Shaping Public Health in a New Europe. Part 2: Evidence Book*, A report to the European Commission. Brussels/Luxembourg: ECSC-EC-EAEC.
- Judge, K., L. Bauld, C. Adams, J. Findlater, A. Killoran, M. Barnes and M. Benzeval (1999) *Health Action Zones: Learning to Make a Difference*. University of Kent at

Evaluation 6(3)

- Canterbury: PSSRU. Full report and executive summary available at <http://www.ukc.ac.uk/PSSRU/ddphazexec.html>
- Kazi, M. and T. May (1999) 'Realistic Evaluation for the Development of Practice', paper presented to UKES Conference, Edinburgh.
- Lumley, J. (1997) 'Preventing Physical and Psychological Health Problems After Birth: The Evolution of PRISM', *Australian Epidemiologist* 4(3): 23–6.
- Macdonald, G., C. Veen and K. Tones (1996) 'Evidence for Success in Health Promotion: Suggestions for Improvement', *Health Education Research* 11(3): 367–76.
- Morgan, A. (1997) 'Experience of Commissioning Reviews', presentation at HEA 'Reviews of Effectiveness Workshop: Their Contribution to Evidence-Based Practice and Purchasing in Health Promotion'. London: RSA.
- National Audit Office (2000) 'Good Practice in Performance Reporting in Executive Agencies and Non-Departmental Public Bodies', report by the Comptroller and Auditor General.
- Nutbeam, D. (1996) 'Health Outcomes and Health Promotion: Defining Success in Health Promotion', *Health Promotion Journal of Australia* 6(2): 58–60.
- Nutbeam, D. (1998) 'Evaluating Health Promotion: Progress, Problems and Solutions', *Health Promotion International* 13: 27–44.
- Nutbeam, D. (1999a) 'The Challenge to Provide "Evidence" in Health Promotion', *Health Promotion International* 14: 99–101.
- Nutbeam, D. (1999b) 'Oakley's Case for Using Randomised Controlled Trials is Misleading', *BMJ* 318: 994 (Letters).
- Oakley, A. (1998a) 'Experimentation and Social Interventions: A Forgotten but Important History', *BMJ* 317: 1239–42.
- Oakley, A. (1998b) 'Experimentation in Social Science: The Case of Health Promotion', *Social Sciences in Health* 4: 73–88.
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: Sage Publications.
- Peersman, G., O. Harden and A. Oakley (1999) *Effectiveness Reviews in Health Promotion*. London: EPI Centre.
- Platt, S. (1998) 'Health Promotion Research: Beyond the Paradigm Wars', keynote presentation at the 1st UK Health Promotion Research Conference, Heriot Watt University, Edinburgh, 6–8 April.
- Potvin, L. (1996) 'Methodological Challenges in Evaluation of Dissemination Programs', *Canadian Journal of Public Health/Revue Canadienne de Santé Publique* 87 Suppl. 2 (Nov.–Dec.): S79–83.
- Quinn Patton, M. (1982) *Practical Evaluation*. Beverley Hills, CA: Sage.
- Ralls, J. and C. Thomson (2000) *The Scotland Index: Visionary Governance for a Small Nation*. Edinburgh: Scottish Council Foundation.
- Sackett, D. L., W. C. Rosenberg, J. A. Muir Gray, R. B. Haynes and W. S. Richardson (1996) 'Evidence Based Medicine: What It Is and What It Isn't', *BMJ* 312: 71–2.
- Sanson-Fisher, R. and E. Campbell (1994) 'Health Research in Australia – Its Role in Achieving the Goals and Targets', *Health Promotion Journal of Australia* 4(3): 28–33.
- Scott, D. and R. Weston (eds) (1998) *Evaluating Health Promotion*. Cheltenham: Stanley Thornes Ltd.
- Scottish Executive (1999) *Towards a Healthier Scotland*, Cmd 4269. Edinburgh: The Stationery Office.
- Speller, V. (1998) 'Quality Assurance Programmes: Their Development and Contribution to Improving Effectiveness in Health Promotion', in D. Scott and R. Weston (eds) *Evaluating Health Promotion*. Cheltenham: Stanley Thornes Ltd.

Wimbush and Watson: Framework for Health Promotion

- Speller, V., A. Learmonth and D. Harrison (1997) 'The Search for Evidence of Effective Health Promotion', *BMJ* 315: 361-3.
- Speller, V., A. Learmonth and D. Harrison (1998a) 'Evaluating Health Promotion is Complex', *BMJ* 316: 1463 (Letters).
- Speller, V., L. Rogers and A. Rushmere (1998b) 'Quality Assessment in Health Promotion Settings', in J. K. Davies and G. Macdonald (eds) *Quality, Evidence and Effectiveness in Health Promotion: Striving for Certainties*. London and New York: Routledge.
- Tones, K. (1996) 'Beyond the Randomised Controlled Trial: A Case for "Judicial Review"', *Health Education Review* 12(2): 1-4.
- Tones, K. (1998) 'Effectiveness in Health Promotion: Indicators and Evidence of Success', in D. Scott and R. Weston (eds) *Evaluating Health Promotion*. Cheltenham: Stanley Thornes Ltd.
- Watson, J. and S. Platt (2000) 'Connecting Policy and Practice: The Challenge for Health Promotion Research', in J. Watson and S. Platt (eds) *Researching Health Promotion*, pp. 1-20. London: Routledge.
- Weiss, C. H. (1995) 'Nothing as Practical as a Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives', in J. P. Connell, A. C. Kubisch, L. B. Schorr and C. H. Weiss (eds) *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*. Washington, DC: The Aspen Institute.
- Weiss, C. H. (1999) 'The Interface Between Evaluation and Public Policy', *Evaluation* 5(4): 468-86.
- Welsh Office (1998) *Better Health Better Wales*. Cardiff: National Assembly for Wales.
- Wiggers, J. and R. Sanson-Fisher (1998) 'Evidence-Based Health Promotion', in D. Scott and R. Weston (eds) *Evaluating Health Promotion*. Cheltenham: Stanley Thornes Ltd.
- Wight, D. (1997) 'Does Sex Education Make a Difference?', *Health Education* 2(March): 52-6.
- Wimbush, E. (1999) 'Strengthening Research Capacity in Health Promotion Practice Settings', *Health Education* 99: 4: 169-76. (<http://www.emerald-library.com/pdfs/14299dd2.pdf>)

ERICA WIMBUSH is Research and Evaluation Manager at the Health Education Board for Scotland (HEBS), responsible for evaluation, external research commissioning and professional development in research and evaluation. Please address correspondence to: Health Education Board for Scotland, Woodburn House, Canaan Lane, Edinburgh EH10 4SG.
[email: erica.wimbush@hebs.scot.nhs.uk]

JONATHAN WATSON is Director of Research and Evaluation at HEBS, responsible for developing and implementing HEBS Research Strategy and the management and planning of HEBS research function. Please address correspondence to: Health Education Board for Scotland, Woodburn House, Canaan Lane, Edinburgh EH10 4SG.
[email: jonathan.watson@hebs.scot.nhs.uk]