Devising the Evaluation Strategy for a Mental Health Training Programme

DI BAILEY

University of Birmingham, UK

ROSEMARY LITTLECHILD

University of Birmingham, UK

What follows is an evaluation of the definitions and methods of evaluation that can be developed into an evaluation strategy. The strategy is derived from the use of a case-study research methodology applied to a UK postgraduate multi-disciplinary Programme in Community Mental Health. This chosen methodology is coupled with a model adapted from the training literature in an attempt to critically review how the strategy might be applied in practice.

KEYWORDS: case study; community mental health; evaluation; evaluation models; multi-disciplinary teams

Introduction

Evaluation is a recurring word in the language of health and welfare agencies in the early 21st century. The changes in social welfare provision in Britain over the past two decades have required service providers to focus on achieving greater cost effectiveness in the delivery of all their services (Department of the Environment, Transport and the Regions, 1998). The British Government have emphasized the need to modernize health and social care and have earmarked funding to support those developments in the education and training of staff which are identified as necessary to underpin service reconfigurations (Department of Health [DOH] 1998a, 1998b).

One way that higher education has been increasingly drawn into the major structural and cultural change required in the entire public sector is through the provision of post-qualifying education and training opportunities for health and social care staff. The government's agenda requires reassurance about the delivery of service orientated training programmes being located within academic institutions, yet there is a dearth of good evaluation of such educational programmes.

The postgraduate Programme in Community Mental Health at Birmingham University, UK is a specific initiative where funding has been provided by the regional National Health Service (NHS) Executive to train mental health workers to deliver contemporary skills-based interventions for people with mental health problems. The programme was approved in 1996 but fits with the vision of contemporary mental health care embodied in the White Paper 'Modernizing Mental Health Services' (DOH, 1998a). It seeks to equip practitioners with the skills, knowledge and attitudes required to deliver practice that is 'safe, sound and supportive' (DOH, 1998a: 3) and is underpinned by interventions which are 'evidence based' and 'outcome driven' (DOH, 1998a: 8).

In this example, the programme of training is designed to 'facilitate learning so that people can become more effective in carrying out aspects of their work' (Bramley, 1996: 2). Yet, according to Georgenson (1982) only about 10 percent of the learning gained from off-the-job courses results in changes in effectiveness at work. So, for social programmes, 'we must be able to think in terms of the linkages between the social problem being addressed' (in this case the care and treatment of people with severe and enduring mental health problem), 'the content and outputs of the interventions strategy' (the training programme) and 'the desired outcomes' (improved knowledge, skills and attitudes of practitioners) (Vanderplaet, 1995: 82). We are, therefore, interested both in the learning that takes place in the academic institution *and* in the worksetting. Work-based learning is encouraged through this particular course design by mental health trusts and local authorities being required to provide work-based supervision and peer support systems for participants.

The Programme in Community Mental Health is multi-disciplinary and participants come from a range of disciplines including nurses, social workers, occupational therapists, clinical psychologists and doctors. Some participants are working as user-advocates or are currently receiving specialist mental health services themselves and therefore able to offer a 'user perspective'. It is a requirement that participants are currently working with people with severe and enduring mental health problems in a paid or voluntary capacity. The programme is run on a part-time, one day per week basis and participants can opt to undertake the one-year Certificate, two-year Diploma or three-year Masters qualification. Places on the programme are currently funded by the West Midlands NHS Executive through training monies 'top sliced' from the mental health trusts.

Devising an evaluation strategy for the programme which is able to span several different domains is therefore important for the following reasons:

- The regional NHS Executive went on to commission an external evaluation of the programme to be conducted over a five-year period by a team of researchers from Durham University. This commenced towards the end of the first academic year of the programme and will focus upon the programme outcomes in relation to any changes in inter-professional working, improved care delivery for service users and the support systems required in the workplace to enable the learning from the programme to be transferred successfully to the practice environment.

- The programme providers, whilst keen to work collaboratively with Durham, need to ensure that the programme meets the requirements for continued academic and professional accreditation.
- The programme providers wish to explore the experience of the training inputs from the learner's perspective.
- Also of interest is the closeness of fit between the training needs hypothesized before setting up the programme, and the extent to which these are representative of the workforce three years on.

In the absence of an overall evaluation strategy, agreed as part of the original tender for the programme, we have endeavoured to develop a co-ordinated framework that is sufficiently robust to serve the programme as it moves into the next three-year cycle of funding in 2000–3. It must be a strategy that is both responsive and proactive, and one that dovetails with the external evaluation process.

This article describes our attempts to devise a retrospective evaluation strategy for this complex programme of learning and development. It will also provide those involved in the design and delivery of interdisciplinary training programmes, that combine both practice and 'academic' components, with an illustration of the complexities that need to be addressed by such an evaluation strategy.

Our paper is first concerned with defining what we mean by evaluation in this context before exploring the methodological issues that faced us in designing the strategy. This will include exploring the roles of ourselves and our colleagues from Durham, UK (with overlapping, but also different perspectives and agendas) and describing the techniques used for collecting and analysing the evaluation data. Our strategy is derived from the evaluation and training literature and reviewed critically in respect of its future application to the Community Mental Health programme in the United Kingdom.

Defining What is Meant by Evaluation

It is important for us to be clear about our definition of evaluation as this will shape the kind of strategy we wish to apply to the programme. A review of the literature reveals that we could adopt any of a number of definitions. It depends upon whether we wish to focus upon the purpose of evaluation or the content, both in terms of what is being evaluated and the processes that it will involve. In respect of these foci there appear in our view to be several distinct strands of argument.

First, that the purpose of evaluation is mainly concerned with establishing the worth of something (Bramley, 1996; Scriven, 1996), be this an established policy or planned intervention (Clarke, 1999), an action or programme (Suchman, 1967) or a total training system or training course (Prior, 1994). In order to do this evaluation, researchers need to be able to collect and provide information about the content, structure and outcomes of programmes (Lincoln and Guba, 1986) for the decision makers involved, which they can use to make decisions about the future of the programme (Norris, 1990).

It is important to establish the worth of the Programme in Community Mental Health, as it is currently delivered within a climate of market testing and we as the providers are in competition with other universities. We also need to be able to provide information to the commissioners of the programme about its worth in terms of value for money and the extent to which it meets identified organizational objectives, changes in work-based performance of participants and any improvements in the standards of care received by service users.

Dissent emerges in the literature in respect of whether the purpose of the evaluation is to prove that certain standards have been achieved (Peryer, 1996) or to improve programme delivery and policy making (Weiss, 1997). The former is often associated with summative evaluation that tends to be conclusion oriented (Patton, 1996) and concerned with the overall effectiveness or impact of the programme, whereas the latter is often associated with formative or action oriented evaluation (Patton, 1996), where the emphasis is on identifying the strengths and weaknesses of the programme.

We would argue that we are concerned with both these functions. As programme providers we perceive two sets of activities associated with our interpretation of evaluation (Pearce, 1995). The first involves issues concerning assessment, validation and examination i.e. the checking of the knowledge, skills and attitudes the learners have acquired against specific academic standards. The second involves training evaluation from a longer term and wider perspective (Prior, 1994) which is concerned with the effect of the training on the groups of learners or their organizations as a whole.

In order to encapsulate these different functions within our definition we quote from Tyler (1950: 69), who was one of the first to define evaluation as a process which links these two elements by:

determining to what extent the educational objectives are actually being realized. . . . However, since educational objectives are essentially changes in human beings. . . . evaluation is the process for determining the degree to which these changes in behaviour are actually taking place.

Given the lack of direction regarding the evaluation methodology from the outset, there is a tendency to set about evaluating the programme in an *ad hoc* fashion, both in respect of how the internal and external evaluations connect, and in the overall collection and analysis of the different sources of data. In order to guard against this, we wish to apply a more formal approach to our evaluation that fits with Lincoln and Guba's 'disciplined inquiry' (1986: 550) which 'applies scientific procedures to the collection and analysis of information about the content, structure and outcomes of programmes' (Clarke, 1999: 1).

We believe that evaluation as a systematic approach is encapsulated by Sander-son's (1995: 124) definition which relates not only to the effects of the training, but also the processes involved. She states that:

Evaluation is the process of obtaining and weighing all the evidence about the effects of training and processes such as identifying training needs or objective setting. It is the systematic collection and analysis of information necessary to make effective decisions related to the selection, adoption, design, modification and value of a training programme.

We therefore wish to adopt this definition of evaluation and explore how we can devise an evaluation strategy to fit within these parameters.

Methods of Evaluation

In order to provide a detailed account of the route we have taken in our evaluation methodology, we will debate the pros and cons of the different approaches to evaluation, associated methods of data collection and the analytical procedures that we have deemed most appropriate. We will also explore the roles of the internal and external evaluators.

Different Approaches to Evaluation

According to McKenzie et al. (1998: 153) 'All authoritative texts on evaluation note that there is no single best method of evaluation.' However, we are already committed to adopting an approach that takes on board both the processes and the effects of the programme within an evaluation strategy that combines both an internal and external stance.

According to Patton (1987: 27) 'formative evaluations often include a process evaluation strategy' in an attempt to describe how the programme actually operates especially in its early stages of development. The primary purpose of such formative evaluation is to identify ways in which the programme can be improved. In the absence of any opportunity to 'pilot' the Programme in Community Mental Health, we recognize the need to obtain ongoing feedback about the programme so that we can make changes. However, we are aware of the need to be able to determine the overall effectiveness of the programme, thus requiring a summative approach. This element of the evaluation strategy is largely being conducted by Durham University in respect of the effects of the training on practice in the worksetting. However, as programme providers, we are concerned with the programme's effectiveness in enabling participants to demonstrate the identified learning outcomes ascribed to each module. We also need to ensure that it meets the standards set by the academic institution and the professional bodies. Thus the distinction between summative and formative approaches often credited to Scriven (1967), whilst relevant, does not embody fully the complexity of the evaluation task at hand.

A review of the evaluation literature revealed that several frameworks have attempted to conceptualize the complexity of evaluations that go beyond this traditional summative-formative dimension. Patton (1982: 44) refers to different categories of evaluation proposed by the Evaluation Research Society that subsume summative approaches under 'impact evaluation' through to 'evaluation of evaluation' which involves evaluators evaluating their own work. Unfortunately, this framework also falls short of describing the programme evaluation being conducted by this study. Two of the stages, front-end analysis and evaluability assessment, are undertaken before a programme is adopted to provide information for those responsible for its planning and implementation. As discussed earlier, one of our difficulties has been designing an evaluation strategy retrospectively.

Chambers et al. (1992) identify a similar framework consisting of five types of evaluation. They range from evaluating programme outcomes, which again subsumes the summative evaluation function but which also incorporates elements of cost effectiveness and cost benefit analysis, through to consumer satisfaction and needs studies. The latter are especially applicable to social programmes that are designed to meet perceived needs within a specific client population. In our example, we need to be able to explore the fit between the training we are providing from the university and the perceived need of the client group. In this case it is for people with severe and enduring mental health problems to receive community oriented services from workers who can deliver effective interventions, including cognitive behavioural approaches, family work and medication management.

Our search for an evaluation approach led us in the direction of the training literature, in addition to the evaluation research literature. Whilst in the training context methods of evaluation are more often referred to as models, we did find evidence of their development towards a series of stages or levels that reflect the frameworks identified above.

Thus, in accordance with Bhola's review of the literature (1998: 329):

Theoretical and methodological developments during the last 40 years show two inter-related tendencies . . . The first is . . . to break away from the classical objectivist outcome based and performance orientated evaluation towards a multiplicity of metaphors and models.

Dickmeyer (1989) refers to metaphors as important concepts in the theory of education research alongside 'model and theory' (p. 151). He regards models as being able to simplify complex systems in order to provide:

a more careful study of the systems operation . . . If we are studying learning . . . we may choose to simplify the situation by using a model . . . to look more closely at isolated aspects of a system, and occasionally, with some flaws, at the system as a whole. (1989: 153)

Dixon (1996) provides a useful summary of evaluation models from the training literature that have traditionally invoked the following four stages.

Level 1 evaluation has tended to focus upon participant reactions at the end of the training programme. This is often known as 'reaction level' evaluation and is obviously short term, taking place immediately after the training and usually conducted internally. Qualitative feedback is captured on evaluation forms often referred to as 'happy sheets'. Whilst this kind of evaluation may inform providers about overall satisfaction with the content of the training, it does little to investigate whether the training needs have been met or whether any learning from the course is being transferred into practice.

Level 2 evaluation is an attempt to take the feedback one stage further to find out whether there have been changes in participants' underpinning knowledge which can be measured by recall and recognition and the application of such knowledge in different situations. Attitudinal shifts can be assessed through changes in behaviour. Usually methods to evaluate at this level include

assignments, observation of practice and collection of evidence to demonstrate the changes in learning.

Level 3 seeks to establish whether there have been any changes in job performance as a result of the training i.e. did the participants apply what they learned in the training on the job? For this reason it is often conducted externally in the worksetting and is usually longer term, relying largely on direct observation and job performance measures.

Level 4 considers the impact of the training upon the organization, particularly whether it has added value to the services that are delivered.

The difficulty with this model lies in establishing whether, or to what extent, any improvements are due to training or to other internal or external factors (Phillips, 1996). It negates much of the 'shaping' work that trainers may be involved in prior to the training event taking place. Whilst we were unable to evaluate this at the time, there were two periods of shaping the Programme in Community Mental Health before its direct delivery commenced. The first one was a period of programme preparation and curriculum design in response to the request for tenders being issued by the regional NHS Executive (NHSE). It was followed by a shorter period of approximately three months to finalize curriculum detail, direct contributions and methods of delivery once the tender had been issued. It is therefore important that any approach to evaluating the programme allows for these efforts to be taken into account.

Other authors have adapted this basic model to include additional aspects. For example, Warr et al. (1970) identify 'context evaluation'. This involves obtaining and using organizational information to identify training needs and objectives. Phillips (1996) separates the evaluation of business results from the question of cost effectiveness. With the pressure on health and social services to improve service delivery, though usually within the availability of current resources, both of these two issues become important (Department of Environment, Transport and the Regions, 1998).

Thus we are faced with the need to identify an evaluation strategy that distinguishes between the functions we need the evaluation to address but at different stages of programme development. Chen (1996) provides a typology that distinguishes between these two elements in an attempt to address the limitations of a summative-formative dichotomy identified earlier. The functions of the evaluation are identified as concerned with either improvement or assessment, combined with the programme stage focused upon either outcome or process.

This approach would allow us to explore whether the Programme in Community Mental Health, as originally conceived by the planners and purchasers, had been implemented successfully as a separate issue to identifying areas for improvement. This approach also attempts to identify which elements are more successful at producing the desired outcomes, thus allowing for the different interests of the internal and external evaluators, together with an overall judgement of the programme in terms of its merit or worth. By transposing Chen's (1996) framework upon a model adapted from the training literature, this provides some useful signposting to the different methods of data collection we need to employ as part of the evaluation strategy.

Shown below (Figure 1) is the schematic representation of the evaluation strategy that we are intending to adopt. This incorporates Warr et al. (1970) stages with the additional outcome of cost effectiveness identified by Chambers et al. (1992) and Phillips (1996) included in stage 4. The model also indicates the orientation of each of the stages in accordance with Chen's (1996) typology in an attempt to be clearer about the different functions of the evaluation process at each stage. Whilst stages 1–3 will rely more heavily on the internal evaluation process conducted by ourselves, stage 4 will predominantly fall in the domain of the Durham researchers. The model aims to show how, in accordance with our definition of evaluation taken from Sanderson (1995), we are adopting a systematic approach to the collection of information at each stage. It also informs our analysis of this information which is critically discussed in the next section.

Data Collection and Analysis

A useful comparison of the methods of data collection, their focus and reporting procedures is provided by Clarke (1999: 8) although this is presented in the context of formative or summative approaches to evaluation. Our research questions and evaluation design require an approach that combines different methods rather than opts for one or the other. Thus, we are persuaded by Guba and Lincoln

Figure 1. Schematic Representations of the Evaluation Strategy Incorporating Stages of the Evaluation Process and Orientation of Evaluation Functions

(1992), who advocate the drawing upon a variety of sources and methods to maximize the chances of considering the influences or contributions from the many factors associated with the education process'. Likewise Cook and Reichardt (1979) urge evaluators to 'adopt a "flexible and adaptive" pragmatic stance' whilst Clarke (1999: 62) exhorts evaluators 'to use whatever methods appear to be the best suited given the nature and context of the evaluation situation'.

Roe and McDonald (1984: 154) describe a wide range of common evaluation methods that are applicable to courses, subjects and teaching in higher education. These include analysis of course documents, feedback from colleagues or collecting student opinions through means of groups discussion (nominal group technique), interviewing or student diaries.

These methods involve collecting both quantitative and qualitative data. The difficulty of utilizing both methods in a combined approach is that they adopt different ontological and epistemological positions. The quantitative paradigm asserts that social research should be 'scientific' and assumes reality is something that is objective. Through this approach, it becomes possible to separate the researcher from the researched using methods such as surveys and experimental designs that employ systematic sampling techniques. In contrast, qualitative approaches seek to provide insight and understanding in order to connect with the views of those involved in the research, thus adopting a more subjective and interactive stance.

Our data collection and analysis draw on both methods and are informed by the research questions our evaluation strategy must address. We have therefore attempted to identify the research questions at each stage of the evaluation and then debate the strengths and weaknesses of the methods of data collection and analysis we intend to use.

Our primary difficulties lie in that we cannot directly observe the causal effects of the Community Mental Health programme but must infer them from the data collected, thereby increasing the possibility of inferential errors. We are also mindful that learners are voluntary participants in the programme and, thus, perhaps more likely than their non-participatory counterparts to be responsive to change their practice and motivated to pursue learning for its intrinsic value. We also recognize that improvement in behaviour is likely to take place as part of the natural process of cognitive development that training promotes. These limitations aside, we believe that there are salient issues that we need to explore as part of the overall evaluation strategy.

1. Context Evaluation

How will we as training providers obtain and use organizational information to measure the training needs and objectives set as part of the programme compared to those identified in the work context?

There are a total of 17 mental health trusts represented on the programme, who are working in various stages of partnership with local authorities and voluntary sector agencies. Each trust is therefore responding to the modernization of mental health services in different contexts. For this reason we need to evaluate the fit between the programme and the strategic focus of the mental health and

Evaluation 7(3)

social care organizations. Data are being collected at a number of levels at the point of recruiting learners from their respective organizations onto the programme. We have designed a formal application process in addition to the standard university application form that allows us to collect *quantitative data* in relation to:

- the number of participants on the programme.
- their respective disciplines and qualifications.
- whether they are from the same or different teams within their organizations or they represent a mix of health, social services and voluntary sector staff.
- the number of years experience they have in working with people with mental health problems.

We are also able to map the discipline of course participants against the discipline of their work-based supervisors. This in itself is providing the focus for a separate article exploring the training needs in respect of inter-professional supervision.

We are also able to detect learners' prioritization of reasons for undertaking the training in respect of the training needs they believe the programme will address in relation to skills-based interventions, improved inter-professional working or applying a user-focused value base to their practice.

We are also able to collect *qualitative data* in relation to the match between the training needs and the programme objectives, from the perspectives of both the employers and the participants. We require employers to provide written justification for why they have nominated particular individuals to undertake the training and how they perceive this selection fits with the overall strategic development of the organization and any training strategy they have in place. Participants are asked to provide a 1000-word reflective account as part of their application, analysing their current training needs and how these reflect practice issues and the need for change in their worksetting.

In their evaluation, Durham are collecting qualitative data through semi-structured interviews with a number of stakeholders including chief executives, middle managers, mental health team members and work-based supervisors. The focus of these interviews will be to what extent the organizational needs help or hinder the training outcomes being realized in the worksetting. In turn, this will allow us, as the training providers, to map this contextual information against the training needs of participants and the programme objectives.

2. Input Evaluation

What training inputs are being offered and how will we explore which are most likely to bring about the desired changes?

The collection of data at this stage overlaps to some degree with stage 3. It again involves a combination of quantitative and qualitative methods including evaluation forms completed at the end of each module of study, together with structured group discussion on evaluation days that take place at the end of each year. The evaluation forms are designed to collect ratings about the quality of the

training inputs for each session on the module, their relevance in relation to the specific topic and the extent to which participants met the learning outcomes of the module.

This allows us to ascertain whether training contributors are using a variety of methods to address the different adult learning styles (Honey and Mumford, 1992) and engage with their audience, as separate from whether the material they deliver is coherent and relevant to contemporary mental health practice. In contrast, open-ended questions on the evaluation form, supplemented by targeted discussion on evaluation days, provide qualitative information about what participants feel they are learning on the course, what inputs they perceive as most valuable and how these relate to specific changes in their practice.

Whilst we accept this is a predominantly subjective approach to evaluating the link between training inputs and desired changes, we acknowledge Richards and Horder's finding (1999) that commitment to collaborative working is fostered by shared personal experience which orients workers away from thinking from an agency or professional perspective.

3. Reaction Evaluation

How will we gauge the participants' reactions to the training?

As might be expected, there are attrition rates with each cohort of learners passing through the programme. Drop-out rates will provide some indirect numerical data on the level of commitment to the training, together with the reasons participants provide as part of the withdrawal process. We have also followed up participants who have left the programme having successfully completed the one-year certificate. Through using semi-structured interviews we have obtained more qualitative data about their experience of the training and their reasons for deciding not to continue.

As outlined above, the evaluation forms completed at the end of each module do ask for qualitative and quantitative feedback about the quality of the teaching, but also seek to establish a subjective assessment from each participant as to what extent they feel the learning objectives identified for each module have been met.

This is a somewhat crude measure as it relies upon self-reporting by the participants themselves. In addition, we recognize that we would not expect the learning objectives for each module to be addressed completely by the taught inputs as practitioners will need some time to try out their learning in their work setting. In order to obtain qualitative feedback on the extent to which this continued learning into practice takes place, structured group discussions will be conducted jointly with both supervisors and participants to revisit the objectives following a set period after module completion. This will allow us to compare whether there are similar views about to what extent the objectives have been met and applied in the ways participants indicated it might be.

4. Outcome Evaluation

Evaluating the impact of the training on mental health service delivery suggests the need for a qualitative methodology with the use of randomized controlled

trials to measure whether explicitly specified objectives are met. This approach would present us with particular ethical difficulties as the programme has been commissioned to be delivered in its agreed format to all participants. We are, therefore, not in a position to allocate learners to control groups who would receive different or no training inputs for particular modules.

However, Miles and Huberman (1984) are of the view that qualitative methods of data collection can also be used to demonstrate the meeting of outcomes. Our difficulty has arisen in developing appropriate indicators to measure to what extent the programme outcomes are achieved and to what extent they impact upon care delivery that can be directly attributed to the programme for reasons mentioned earlier.

However, by adopting an evaluation approach that incorporates an analysis of cost effectiveness issues, we acknowledge that evaluation of outcome measures needs to connect with resource issues. This is a particularly important issue for the employers of mental health workers and the programme purchasers who need to be aware of the 'cost' of achieving change in mental health practice.

We have attempted to measure the outcomes of the training in terms of:

Skills, Knowledge and Attitudes that have Changed Data are being collected at different levels using different methodologies to reflect the different types of outcomes in respect of academic and professional standards, work-based outcomes and the outcomes for people using mental health services.

Skill delivery is an important part of the programme as it is expected that participants will leave the course being proficient in delivering family work and cognitive behavioural interventions. These are currently assessed through assignments at the end of each module. However, we recognize that participants' ability to describe a skill is not the same as demonstrating that skill in practice. We have, therefore, put in place a tutoring scheme in which all participants will engage in the workplace, in an attempt to connect the teaching of skills to their use in practice. This will also allow for direct observation of skill development. Participants will meet in small groups of no more than three with a tutor, to practise role playing of client scenarios and problem solve any issues in implementing the approach.

A weakness in this approach is that we do not currently have a means of observing skill delivery prior to participants commencing the programme, from which to establish a baseline and use either a pre-test post-test design or static group comparison (Campbell and Stanley, 1963). Thus, this is one area where we will need to make causal inferences against the learning outcomes of the skills-based modules and are limited to the post-test only design which according to Clarke (1999) is the weakest of the three types of quasi-experimental designs.

In respect of a different outcome measure, written assignments are required at the end of each module and take different forms depending upon whether their focus is on skills, knowledge or values. All are rooted in work-based practice and use a variety of techniques to demonstrate that the learning outcomes for the module have been met. They include case studies, project work, audio taped interviews of sessions of direct family work and portfolios of evidence from the participants' work. The assignments provide a source of quantitative data in

terms of how many participants are meeting the academic and professional standards at which the course is set, and we are able to analyse this further in respect of any relationship with a participant's discipline and the level of active supervision being received in the worksetting.

However, we recognize that detecting changes in knowledge, skills and attitudes, as demonstrated through the assignments, poses a similar degree of difficulty. This is because practice written about is not always the same as practice delivered, which can only be realistically checked out through direct observation against specific outcomes. Whilst the assignments may demonstrate practitioners' ability to integrate literature, values and self-reflection into their thinking, the providers of the programme are making an assumption that such thinking directly translates to practice.

However, our colleagues from Durham have been able to implement a more quantitative outcome-based design into their evaluation in two respects. First, they are collecting attitudinal measures, using validated instruments in respect of multi-disciplinary working at yearly intervals. These are being compared with groups of learners on other training programmes that do not have a multidisciplinary focus. Whilst this provides an example of post-test only non-equivalent groups design (Mark and Cook, 1984), we need to recognize that the course content and training methods employed will differ generally between the different programmes, thus limiting to what extent any attitudinal change can be attributed directly to the multi-disciplinary focus of the Birmingham course.

Participants' Job Performance Outcomes in relation to job performance will also be measured by the Durham team through qualitative open-ended interviews with participants and managers in the trusts and local authorities. They will seek to identify the variables required in the workplace to enable practitioners to transfer their learning into the work setting effectively and over a significant time period.

Using the same instruments identified above, the Durham team are also exploring improvements in team working and clarity about professional roles that are again being compared with groups of workers on other training programmes that do not have an inter-professional focus. It is hoped that these data will begin to offer some clues as to whether the particular focus on inter-professional working in the training delivery is being reflected in improved inter-professional relationships outside the training room.

Improved Job Performance being Translated into Better Services for Users In accordance with a more quantitative methodology, the Durham team have also systematically sampled the caseloads of participants on the course and are interviewing a random sample of service users. This is a particular illustration of where the researchers are separate from the researched as users of mental health services in other parts of the country are being trained to conduct the interviews.

This will provide qualitative data on any perceived changes in practitioner competence from the users' perspective. Unfortunately this aspect of the evaluation does not go as far as identifying any objective indicators of improved outcomes

(such as admission rates to hospital, or exit rates from services) that can be incorporated into a more comprehensive measure of training effectiveness.

Cost Effectiveness of the Programme The cost effectiveness of the programme is important to a number of stakeholders, particularly the purchasers but also the employers, as the funding for the training is centrally top sliced from trusts' training budgets.

It is possible to devise fairly simplistic measures of cost effectiveness, as the programme providers are able to analyse the amount of funding being provided compared to success in student numbers achieving the qualifications associated with the programme. Much more difficult is measuring the impact of indirect costs, such as time out from the workplace to attend the course and time taken in work-based supervision and to practise the skills learned. In order to begin to explore these elements a series of semi-structured interviews will be undertaken by the Durham team with Chief Executives of the trusts to consult on the continued 'top slicing' of education and training money to fund the programme. This is one area where the chosen methodology is limited. As we are not looking at the costs of other programmes, there can be no comparison of the cost-effectiveness of delivering the training in this way.

Role of Evaluators

Our approaches to data collection and analysis described above indicate a complex relationship between ourselves, as programme providers engaged in the process of evaluation, and our colleagues from Durham. They were commissioned to undertake an external evaluation, but have chosen to work closely with us in the interests of ongoing programme development. Any analysis of the roles of the evaluators therefore has to take into account the complexities of these relationships.

Our colleagues from Durham were commissioned as 'independent researchers' to undertake an evaluation of the programme on behalf of the regional NHSE. They have no long-term role within the programme (Patton, 1986). However, in the short term, as part of the overall data collection strategy, they have become increasingly involved in teaching students how to collect the required outcome-based measures. Furthermore, they have given feedback to the students about the attitudinal measures in relation to inter-professional working. Another aspect of teaching has been describing the policy framework underpinning the study.

The Durham researchers have also participated in the structured group discussion on evaluation days as participant observers. In doing so they have identified with the aims of the regional NHSE and programme providers in delivering better educational programmes (Mathison, 1994). However, given the limitations of the methodology of the internal evaluations described above, they have been able to provide an overview of other training initiatives. These have served as comparisons to our own evaluation and have influenced the use of a wider range of evaluation techniques (Love, 1991).

The nature of the collaborative relationship that has developed between

ourselves and Durham has allowed us to guard against most of the disadvantages of an external evaluation (Clarke, 1999: 23) and capitalize on the positive aspects.

The role of the evaluators in this study has also taken on another dimension with the collaboration extending to the organizations and stakeholders involved with the programme. According to Cronbach et al. (1980), the evaluator should act as both the informant and educator to all parties in the evaluation and this involves making information available to all the groups interested in the programme. In our example this includes employers, managers, current and potential participants and people who are using the services being affected by the training.

Working in a responsive mode (Stake, 1995), we, as the internal evaluators, and our colleagues from Durham have direct contact with participants, their supervisors and managers in the local authorities and trusts. We have begun to explore both their different value positions, practical problems with respect to the programme, the difficulties in implementing change in mental health provision and the part training has to play in improving services. We would argue that through the process of collaborating with this range of stakeholders, we have attempted to develop an evaluation process that is simultaneously 'diagnostic, change oriented and educative for all parties involved' (Lincoln and Guba, 1985: 141).

In accordance with empowerment evaluation (Fetterman, 1994), we are using this evaluation to indicate to employers the extent to which the training can act as a catalyst to change in practice. Equally important is the role employers need to play in influencing service structures and caseload management to ensure that they help themselves to get the maximum benefit of the training initiative.

Conclusion

Our approach to designing an evaluation strategy reflects a number of tensions and diversity of views in the literature with respect to definitions of evaluation, methods and models together with different approaches to the collection and analysis of data. As programme providers and internal evaluators of the Programme in Community Mental Health, we are satisfied that the strategy we have outlined will provide us with sufficient information to learn about the programme we are delivering. The methods adopted will enable us to both improve the programme and begin to assess whether we are achieving the desired outcomes. By transposing Chen's (1996) typology onto our stage model, we should also be clearer about the functions of the evaluation process at each stage.

In accordance with Patton (1986), we would argue that the Programme in Community Mental Health has provided us with an opportunity for combining an internal and external evaluation in one overall strategy. Thus, we are able to benefit from the advantages each one has to offer. By working collaboratively with our colleagues from Durham, we are attempting to combine both qualitative and quantitative methods of data collection and analysis, and influence the traditional roles of internal and external evaluators in the best interests of programme development and collaboration with employers.

We accept that we are describing an evaluation strategy which is complex and will be time consuming to conduct. Referring back to Vanderplaat's (1995) work,

this simply reflects the complex nature of the relationship between the social problem with which we are grappling, that of improving the care and treatment for people with severe and enduring mental health problems, a relevant training programme and the desired outcome of improved knowledge, skills and attitudes of practitioners.

The very act of evaluation has led us to a deeper understanding of the wealth of material and evidence to be generated and the range of evaluation methods and data collection strategies we could employ. Thus, in attempting to generalize from our particular example, we would suggest that academic providers of post-registration education and training programmes go beyond the standard internal evaluation requirements of academic institutions. They need to utilize the wealth of evaluation research and literature to devise an evaluation strategy that in itself adds to the worth of the training being delivered. They also need to ensure that their strategy adopts tried and qualitative and quantitative measures, and where possible transcends the boundaries of internal and external evaluation in the best interests of programme development and stakeholder involvement.

References

- Bhola, H. S. (1998) 'Constructivist Capacity Building and Systemic Evaluation of Adult Basic Education and Training in South Africa', *Evaluation* 4(3) 329–50.
- Bramley, P. (1996) *Evaluating Training*. London: Institute of Personnel and Development.
- Campbell, D. T. and J. C. Stanley (1963) *Experimental and Quasi-experimental Design for Research*. Chicago, IL: Rand McNally and Company.
- Chambers, D. E., K. R. Wedel and M. K. Rodwell (1992) *Evaluating Social Programmes*. Boston, MA: Allyn and Bacon.
- Chen, H. T. (1996) 'A Comprehensive Typology for Programme Evaluation', *Evaluation Practice* 17(2): 121–30.
- Clarke, A. (1999) *Evaluation Research: An Introduction to Principles, Methods and Practice*. London: Sage.
- Cook, T. D. and C. S. Reichardt (1979) *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, CA: Sage.
- Cronbach, L. and Associates (1980) 'Toward Reform of Programme Evaluation', in R. Stake (1995) *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Department of the Environment, Transport and the Regions (1998) *Modernizing Local Government*. London: HMSO.
- Department of Health (1998a) *Modernizing Mental Health Services: Safe, Sound and Supportive*. London: HMSO.
- Department of Health (1998b) *Modernizing Social Services*, Government White Paper Cm 4169. London: HMSO.
- Dickmeyer, N. (1989) 'Metaphor, Model and Theory in Education Research', *Teachers College Record* 21(2): 151–60.
- Dixon, N. M. (1996) 'New Routes to Evaluation', *Training and Development Journal* May: 82–5.
- Fetterman, D. M. (1994) 'Empowerment Evaluation', *Evaluation Practice* 15(1): 1–15.
- Georgenson, D. L. (1982) 'The Problem of Transfer Calls for Partnership', *Training and Development Journal* 36(10): 75–8.

Bailey & Littlechild: Evaluation Strategy for a Mental Health Training Programme

- Guba, E. G. and Y. S. Lincoln (1992) *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsiveness and Naturalistic Approaches*. San Francisco, CA: Jossey-Bass.
- Honey, P. and A. Mumford (1992) *Manual of Learning Styles*. Maidenhead: Honey.
- Lincoln, Y. S. and E. G. Guba (1985) *Naturalistic Inquiry*. Newbury Park, CA: Sage.
- Lincoln, Y. S. and E. G. Guba (1986) 'Research, Evaluation and Policy Analysis: Heuristics for Disciplined Inquiry', *Policy Studies Review* 5: 546–65.
- Love, A. (1991) *Internal Evaluation*. London: Sage.
- McKenzie, J., S. Sheely and K. Twigwell (1998) 'Drawing on Experience: An Holistic Approach to Student Evaluation of Courses', *Assessment and Evaluation in Higher Education* 23(2): 153–63.
- Mark, M. M. and T. D. Cook (1984) 'Design of Randomized Experiments and Quasi Experiments' in L. Rutman (ed.) *Evaluation Research Methods: A Basic Guide*, 2nd edn. Beverly Hills, CA: Sage.
- Mathison, S. (1994) 'Rethinking the Evaluator Role: Partnerships between Organization and Evaluators', *Evaluation and Programme Planning* 17(3): 299–304.
- Miles, M. B. and A. M. Huberman (1984) *Qualitative Data Analysis*. London: Sage.
- Norris, N. (1990) *Understanding Educational Evaluation*. London: Kogan Page.
- Patton, M. Q. (1982) *Practical Evaluation*. Beverly Hills, CA: Sage.
- Patton, M. Q. (1986) *Utilization-focused Evaluation*, 2nd edn. Newbury Park, CA: Sage.
- Patton, M. Q. (1987) *How to use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage.
- Patton, M. Q. (1996) 'A World Larger than Formative and Summative', *Evaluation Practice* 17(2): 131–44.
- Pearce, I. D. (1995) *The Assessment and Evaluation of Training: A Practical Guide*. Hitchin: Technical Communications Publishing Ltd.
- Peryer, A. (1996) in Department of Health (1997) *Get Going: A Guide to the Evaluation of Training*. Milton Keynes: Joint Initiative for Community Care.
- Phillips, J. (1996) 'Was it the Training?', *Training and Development Journal* March: 28–32.
- Prior, J. (ed.) (1994) *Gower Handbook of Training and Development*, 2nd edn. Aldershot: Gower.
- Richards, G. and W. Horder (1999) 'Mental Health Training: the Process of Collaboration', *Social Work Education* 18(4): 449–58.
- Roe, E. and R. McDonald (1984) 'Informed Professional Judgement', in J. McKenzie, S. Sheely and K. Twigwell (1998) 'Drawing on Experience: An Holistic Approach to Student Evaluation of Courses', *Assessment and Evaluation in Higher Education* 23(2): 153–63.
- Sanderson, G. (1995) 'Objective and Evaluation', in S. Truelove (ed.) *The Handbook of Training and Development*, 2nd edn. Oxford: Blackwell.
- Scriven, M. (1967) 'The Methodology of Evaluation', in A. Clarke (1999) *Evaluation Research: An Introduction to Principles, Methods and Practice*. London: Sage.
- Scriven, M. (1996) 'The Theory Behind Practical Evaluation', *Evaluation* 2(4): 393–404.
- Stake, R. E. (1995) *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Suchman, E. A. (1967) *Evaluative Research*. New York: Russell Sage.
- Tyler, R. W. (1950) 'Basic Principles of Curriculum and Instruction', in E. G. Guba and Y. S. Lincoln (1992) *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsiveness and Naturalistic Approaches*. San Francisco, CA: Jossey-Bass.
- Vanderplaat, M. (1995) 'Beyond Technique: Issues in Evaluating for Empowerment', *Evaluation* 1(1): 81–96.
- Warr, P. and M. Bird (1970) *Evaluation of Management Training*. Aldershot: Gower.

Evaluation 7(3)

Weiss, C. H. (1997) 'How can Theory Based Evaluation Make Greater Headway?', *Evaluation Review* 21(4): 501–24.

DI BAILEY is the Co-director of the Interdisciplinary Centre for Mental Health at the University of Birmingham and the Director for a multi-disciplinary Postgraduate Programme in Community Mental Health. Please address correspondence to: Department of Social Policy and Social Work, University of Birmingham, Edgbaston, Birmingham B15 2TT.
[email: d.e.bailey@bham.ac.uk]

ROSEMARY LITTLECHILD is a lecturer in Social Work. Her teaching research and publication interests are in community care and staff training and development. Please address correspondence to: Department of Social Policy and Social Work, University of Birmingham, Edgbaston, Birmingham B15 2TT.