Connecting policy sectors in evaluation: challenges of evaluating sustainable development

Bart De Peuter
Researcher
Catholic University of Leuven, Belgium
Email: bart.depeuter@soc.kuleuven.be

Co author: Joris De Smedt

Sustainable development receives increasingly attention from policy-makers. The concept stresses the need for a holistic approach by which ecological as well as social and economic concerns are to be integrated within the strategic policy and program objectives and related decision-making. However, these three policy fields, being considered as the pillars of sustainability, also reflect different sets of values to be balanced.

Connecting and integrating the three policy areas in an evaluation from a joint sustainability perspective entails several challenges for each stage of the evaluation process: design, observation, analysis and judgment. In the design stage the evaluation is structured by determining its focus through the selection of evaluative questions, criteria and techniques. Here, balancing between the different policy perspectives and their sets of values becomes an important but challenging task.

As to observation and analysis the complexity of the evaluandum brings along a large amount of data to be gathered and processed. Together with the longer time horizon of sustainable development, it will not always be easy to obtain reliable and valid measurements.

Connecting policy sectors in evaluation also entails methodological issues and choices for analysis and judgment. One option is a single method approach directly integrating the policy perspectives. Another possibility is a multi-method approach using a different evaluation technique for each policy sector involved. However, the question arises whether and how an overall or synthetic judgment can be reached.

This paper will map out these challenges and identify critical success factors to connect different policy sectors from a common evaluation perspective. The evaluation of the strategic traffic mobility plan of Antwerp will be used as a case example. The ex ante evaluation of this plan consisted of a strategic environmental assessment, a cost benefit analysis and a social impact analysis, reflecting the three pillars of sustainable development.
'Measuring the tail of the dog that doesn't bark in the night' - challenges in evaluating the Scottish suicide prevention strategy

Mhairi Mackenzie
Lecturer
University of Glasgow, UK
Email: m.mackenzie@clinmed.gla.ac.uk

Presentation on behalf of Choose Life Research Consortium: Mhairi Mackenzie, Emma Halliday, Allyson McCollam, David McDaid, , Joanne MacLean, Stephen Platt, Amy Woodhouse.

The Scottish suicide prevention strategy 'Choose Life' was launched in December 2002 and a national evaluation was commissioned from a range of collaborating institutions in 2004. This paper discusses the range of challenges faced in evaluating such a programme and in influencing subsequent government policy. These include the following: tracing the impact of current interventions on the prevention of outcomes in the long distant future (for example, how do interventions in schools impact on suicide prevalence?); linking processes, outputs and intended outcomes using a theory-based approach (for example, how convincing is it that a training programme reaching a particular target audience will contribute to a 20% reduction in suicide?); providing an economic perspective in the absence of outcome data (for example, is it meaningful to say that a programme has the potential to be cost-effective?); providing a framework for future attribution of change in suicide incidence to the existence of the national programme (for example, when suicide rates are subject to major year-on-year fluctuation at area level).

The research consortium invites discussion of these issues and will share its views on the most productive way forward for such evaluations.
Defeating Douglas Adams - Developing a cross-cutting approach to evaluating the impact of Regional Development Agencies

Carol Candler
Evaluation & Best Practice Manager
One NorthEast, UK
Email: carol.candler@onenortheast.co.uk

*Human beings, who are almost unique in having the ability to learn from the experience of others, are also remarkable for their apparent disinclination to do so.*

Douglas Adams, "Last Chance to See"

Public sector bodies are not renowned for their openness and ability to learn from experience, stemming partly from a constant need to justify the value for the public money they spend. Regional Development Agencies are currently under pressure to demonstrate their impact and value as part of a comprehensive spending review and are wrestling with the difficulty of producing evaluation evidence that will justify their existence without giving ammunition to their opponents. This presents the thorny question of whose interests such evaluations serve.

The RDA’s are therefore facing the challenge of how to encourage effective and independent evaluation in a culture where nothing can ‘fail’. Alongside this they need to balance the need for independence in evaluation with the pressure to cut costs. This paper outlines the approach taken by One NorthEast to establish a systematic approach to evaluation that will provide a coherent picture of the impact of its interventions, whilst promoting the use of ‘learning from experience’ at all levels.

Recent work in ONE, including a local review of Area Based Interventions (IDOX, 2006) and an Audit of ONE Evaluations undertaken as part of the implementation of the DTI Impact Evaluation Framework (Watson, 2006) has highlighted gaps and inconsistencies in evaluation evidence for RDA interventions. In One NorthEast the response has been to develop a tiered approach to evaluation, supporting project, programme and policy evaluation. Key to this approach has been the development of a framework for evaluations that includes work at the intersections between these levels, evaluating the ‘added value’ of programmes over individual projects and of the Agency over individual programmes.

The approach is a work-in-progress but the paper will outline progress to date and plans to ‘evaluate the evaluation’.
The challenge of evaluating a city growth strategy initiative

Susan Bagwell
Research Development Manager
Cities Research Institute, London Metropolitan University, UK
Email: s.bagwell@londonmet.ac.uk

Business clusters are viewed by policy makers world-wide as an important vehicle for economic development yet the evaluation of cluster policy has been acknowledged as being undeveloped (Andersson et al 2004, Dti/Ecotec 2004). More recently clusters have been promoted as a means of encouraging the regeneration of deprived inner city areas (Porter 1998) and this US inspired model of business led regeneration has led to the introduction of the City Growth Strategy initiative in the UK. This paper explores the challenges in developing an evaluation framework for the formative evaluation of one of the City Growth Strategy (CGS) pilots based in the City Fringe area of north London. The paper highlights the issues involved in trying to evaluate a complex Area Based Initiative such as the CGS which involves multiple objectives, levels, timeframes and funders, individual agents (businesses) and their interaction (business clusters), and which poses particular challenges in identifying causal relationships between project and business level outputs and the broader social-economic outcomes the initiative is trying to achieve. The need to address the diverse goals of business and public sector stakeholders adds a further dimension.

The paper proposes recommendations for changes to evaluation practice including a critical reflection on performance indicators, methods for complexity, capacity building for regeneration practitioners, and the use of case studies. Finally, drawing on this on-going practical experience, the paper will conclude by contributing to the debate on the evaluation of regeneration policy.
A Swedish trade union's prerequisites to administer a national working environment program

Dr Johan Hansson
Researcher/evaluator
Umeå Centre for Evaluation Research
Umeå University, Sweden
Email: Johan.Hansson@ucer.umu.se

In 2005, a large Swedish employee trade union in the government sector initiated a national working environment program. The program aimed at reducing work related ill-health. Through training efforts, designed to provide the participants with knowledge and skills needed for measures at the workplace, the expected outcome was assumed to be achieved. 600 union members from eight government authorities, all employed at superior management levels, participated in the program activities.

To attain scientific legitimacy, a university evaluation centre was assigned to evaluate the program. Joint efforts between practitioners and evaluators/researchers were organized in reference group meetings. A quasi-experimental evaluation approach, including a subobjective design, was applied to determine the extent to which the program had attained the projected outcomes of decreased workplace-related ill-health.

A central question examined in this paper concerns the trade unions specific prerequisites to utilize the evaluation results to administer and improve the original program. The role of the evaluator is revised in relation to the trade unions political interests, the program sponsors financial concerns and finally, from a client/employee perspective.

A general discussion about the trade unions prerequisites to manage organizational change activities in the government sector and in state companies is also completed. Ethical and methodological implications from the stakeholder groups’ diverse priorities are considered as well.
A bird's eye view of multiple-indicators reports

Dr Rita Sever
Senior Lecturer and Researcher
The Hebrew University of Jerusalem, Israel
Email: sela29@bezeqint.net

Co author
Ilan Shina
Evaluation and Research Unit, Israeli Karev Educational Program
Email: ilanshin@012.net.il

There's a growing demand for evidence-based policy making and accountability. High stakes decision making requires highly complex information; multiple indicators, combinations of quantitative and qualitative data - since important evidence is often collected qualitatively rather then quantitatively.

Thus policy makers are served with (or flooded by) evaluation reports which contain a large number of indicators that are measured on a variety of non-equivalent scales. E.g. standardized test scores together with ordinal attitude scales, measures of organizational climate etc.

This abundance of data may often bring about a boomerang-effect. If information becomes too complex for the human mind to process systematically (via rational, cybernetic models as recommended by decision making theories), it may be processed via more intuitive, partial, uncontrolled mechanisms - namely the "automatic pilot" of the mind.

Therefore policy makers need parsimonious yet comprehensive frameworks for systematically organizing large quantities of highly diversified information.

In response to this need, a generic methodological framework has been developed, which helps organize multiple-indicators reports in a way that enables policy makers to cope with them effectively. For purposes such as comparing at one time the general status of several systems (such as schools, or municipal education systems), or building a time-series to monitor changes from one year to the following (or previous) years.

The main features of the framework are :
1. It's generic (applicable to many different sets of indicators)
2. orthogonal rotation: indicators become "observations"
3. definition of a common point of reference - (e.g. benchmark; national/state/ district average; result on first year's measurement)
4. transformation : one scale for all "observations"- each "observation" is scored on the same scale: above/ within/below the chosen point of reference
5. non-parametric approach, e.g. distribution of all "observations" (i.e. indicators) on the above 3-point scale
6. Zoom-in, Zoom out - between a bird's eye view of the forest and a close look at a single tree.
   Between the most parsimonious, compressed presentation of data, down to the most specific, qualitative information.

The paper demonstrates the framework's usefulness by presenting its application for analyzing an external multiple-indicators evaluation report (MEITZAV = indicators of school effectiveness) which educational policy-makers in Israel receive on a yearly basis.
Combining biometric and self-report measures of smoking

Professor Melinda Davis  
Research Assistant Professor  
Evaluation Group for Analysis of Data  
Department of Pediatrics  
University of Arizona Health Sciences Center, USA  
Email: mfd@email.arizona.edu

Biometric measures are considered the gold standard for assessing levels of smoking and the accuracy of self-report measures is usually regarded as biased. However, physiological measures are usually limited to recent smoking behavior and their cost may be prohibitive, while self-report measures can be inexpensively collected through a variety of methods.

Biochemical and self-report data have been collected in two samples that differ in important ways and are expected to have different sources of biases. Structural equation modeling was used to combine the multiple physical and self-report measures, and the results provide information about the relative merits of these indicators within and across studies.
It’s hard to help young people to give up smoking – and so is producing rigorous, conclusive evaluation for evidence-informed policy making

Dr Wendy Gnich
Research Unit on Health Behaviour & Change
University of Edinburgh Medical School, UK
Email: wendy.gnich@ed.ac.uk

Mark Bitel
Partners in Evaluation Scotland, UK
Email: mark@evaluation.u-net.com

Reducing the prevalence of smoking is critical to improving the health of populations and developing interventions to help people to stop smoking is the foremost public health improvement approach to reducing cancer and other preventable cardiovascular diseases. The role of smoking cessation services is even more critical given the policy trends to ban smoking in enclosed public places, such as the new laws introduced in Scotland in March 2006.

This paper reports on the findings of a four-year study of a pilot programme across Scotland, ending in March 2006, to develop the evidence base about smoking cessation interventions in a variety of settings to help young people stop smoking. These included a maternity service, schools and colleges, youth clubs, and an on-line chatroom.

The evaluation employed rigorous quantitative outcome measurement, supported by a qualitative case study approach following the progress of each of the eight settings. In addition, each project was responsible for their own self-evaluation, as well as participating in the programme-wide external evaluation.

The paper shows how the qualitative data enriched the quantitative outcome data. It discusses the implications for quantitative evaluation designs that need high numbers to produce powerful statistical findings and the usefulness of data with lower than expected numbers. It also highlights the discrepancies between the results of the individual project self-evaluations and the results of the external programme evaluation. Finally, it presents the findings and their implications for public health approaches to reduce smoking prevalence in Scotland.
The National Evaluation of the Children's Fund: Learning for Practice and Policy

Paul Mason
Institute of Applied Social Studies
University of Birmingham, UK
Email: mailto:p.mason.1@bham.ac.uk

Ian Plewis
Institute of Education
University of London, UK

Nathan Hughes
Institute of Applied Social Studies
University of Birmingham, UK

Paper One
Nathan Hughes – National Evaluation of the Children's Fund: Qualitative Research for Policy and Practice

This paper will provide an overview of the National Evaluation of the Children's Fund: how the aims and design sought to ensure that learning was fed into policy and practice as it developed at local, regional and national level, and how this was achieved. Examples will include learning about local partnership working, learning about work with marginalised groups, and how this was utilised.

Paper Two

This paper will describe and discuss how NECF sought to understand the impact of Children's Fund services. Work with existing data sources to inform multi-level modelling reduced the burden upon participants and services. Yet issues with data availability, and the time frame of the evaluation, pose challenges in understanding impact. Policy makers and practitioners can struggle concepts concerning estimates of impact and the issues that limit our understandings.

Paper Three

This paper will discuss tensions in the relationship between a large complex evaluation and policy and practice stakeholders. Commissioned to a complex design, changing structures for evaluation governance, and major changes to the policy and the broader policy context, posed challenges for the evaluation and created tensions with policy and practice stakeholders.
Who is afraid of the EU evaluation?

Monika Mura
PhD Student
Leeds Metropolitan University, UK
Email: mn_k@yahoo.com

This paper addresses the following question: what is the reaction to the evaluation, both as an activity and as an aid to decision?

The EU Commission, in the meta-evaluation report issued in November 2004 affirms that the evaluation capacity in EU member states that benefit from the Structural Funds aid has considerably increased if compared with the past experience.

More space is given to independent evaluators in the fora with social partners and with dedicated sessions of the monitoring committees, while ad hoc institutions have been created to better deal with the evaluation.

If this is true, evaluation should really play a central role in the reprogramming of objectives, strategies and objectives.

However, a deeper analysis of the organisations which currently undertake the evaluation and of the relationship between evaluators and users -which are programme managers and the policy-makers, might demonstrate that there is still a long way to cover before reaching the goal of an actually grown evaluation capacity.

Starting from the analysis of the evaluation as an aid to decision and policy-making, and the function that it should have, this paper will examine how evaluators and commissioners react to the evaluation.

The former perhaps consider it as a business like all the others, and act with the only purpose to obtain future assignments from the clients. The latter might see the evaluation as a threat to their position or an intrusion to their policy-making power, for which they wish to have an exclusive domain.

Cases studies are drawn from Italy and Ireland, as they experienced a different performance in the use of the EU Structural Funds, and adopt a different approach to evaluation. The issues that will be investigated cover the analysis of their evaluation reports, their evaluators and the relationship among the latter and their commissioners.
The Evolution of Evaluation in the EU

Francis McGowan
Lecturer
Sussex European Institute, University of Sussex, UK
Email: f.mcgowan@sussex.ac.uk

This paper explores the way in which evaluation has become a core part of EU policy-making (and in the process bring together the study of European integration and the study of Policy Evaluation). Over the last 10 years there has been a rapid dissemination of evaluation techniques from areas where budgetary questions are prominent to the whole range of EU policies. Moreover the techniques themselves and the purposes of evaluation have undergone a significant evolution, becoming an important component of policy making. Yet the meta-dimension of this process has been relatively underresearched (certainly compared to other aspects of EU policymaking such as agenda setting, institutional bargaining and implementation). The EU institutions expend considerable resources on evaluation, academics and others evaluate particular areas of EU policy but there is less coverage of the impact of evaluation per se.

This paper seeks to examine how evaluation has been used within the EU, mapping the history of its development and distinguishing between the different dimensions of evaluation. In looking at the role of evaluation in the EU, the paper will draw upon the political science literature on "policy learning" and the European integration literature on policy making in the EU. The former strand will enable us to examine how far evaluation helps to facilitate learning in terms of the conduct of policy in particular areas and how far there is policy learning within the evaluation community inside the EU institutions. The latter enables us to consider how far evaluation contributes to the integration process.
Use of evaluation in the domain of agriculture and rural development

Martin Scheele
European Commission
DG Agriculture and Rural Development, Belgium

The Common Agricultural Policy is a matter of Community policy competence, which implies that the foundations of this policy are developed by the European Commission, with its specific rhythm set by the cycle of the Financial Perspective and commitments laid down in legislation and Council conclusions. In this respect, Directorate General for Agriculture and Rural Development has a fully developed analytical capacity. However, external evaluation, carried out regularly in line with the requirements of the Financial Regulation, can be valuable source of assessment and judgement. In order to be useful, evaluation results must be available timely, with sufficiently precise answers to evaluations questions based on an appropriate balance of quantitative and qualitative assessment, judgements that is solidly based on findings, and a language that can be understood.

The presentation will provide insights from a decade of experience with using evaluation results within the Directorate General for Agriculture and Rural Development. It will elaborate the criteria to be met, the limitations faced, and examples of success. These matters will be discussed with respect to both the evaluation management by the Commission and the work carried out by the evaluators. Particular emphasis will be given to the aspects of planning and preparation of evaluation projects, the limitations resulting from procurement procedures with respect to responding to new policy events at short notice, the availability of data and analytical methods, the conditions to be met for ensuring that evaluation results are solid and defendable, the involvement of different stakeholders, and the importance of communication skills on the side of evaluators.

As a special case, the presentation will refer to the arrangements made concerning the evaluation of Rural Development Policies, where the Commission has the role of providing guidelines and continuously co-ordinating evaluations carried out by Member States.
The European project and EU evaluation

Professor Nicoletta Stame
University of Roma "La Sapienza", Italy
Email: nstame@aconet.it

Professor Claudio M. Radaelli
Exeter University, UK
Email: c.radaelli@exeter.ac.uk

Professor Jean-Claude Barbier
CNRS and University of Paris, France
Email: Jean-Claude.Barbier@univ-paris1.fr

The European project, the EU crisis and evaluation
Nicolella Stame
Following the No vote on the Constitution, the EU institutions have been unable to react and choose the opportunity for rethinking their role, and reframe the main political strategies, such as the Lisbon strategy. The European citizens' uneasiness revealed by that vote touches upon various aspects, ranging from the aloofness of institutions to the fear of losing national standards in the welfare state.

Could evaluation at the EU level contribute to offer an understanding of those predicaments, or has EU evaluation been part of the problem? Evaluation is a tool for transparency, accountability and learning that has been granted many European states by the EU tools, like programs and Structural funds: as such it should have helped the dialogue between the Commission and the citizens. But EU evaluation methods and procedures have been conceived inside the very institutional setting that is under attack: the "inspection tradition" has resisted side by side with other trends, such as the "good practices" presented by the Open Method of Coordination, and there has been a scarce interaction with the advancements in evaluation approaches.

A new season of dialogue inside the European evaluation community may open the way to a positive role played by EU evaluation in providing new insights into both controversial policy issues (like the internal market for services, or the agricultural policy) and institutional issues (like the subsidiarity principle).

Whither "better regulation" policy for the Lisbon agenda?
Claudio Radaelli
The paper critically assesses the role played by the initiatives for regulatory quality (also known as 'better regulation' policy) in the context of the Lisbon agenda. It provides a re-conceptualization of better regulation as meta-regulation, to be analyzed in its structural and discursive properties. At the EU level, there are some features of an emerging open method of coordination for better regulation, and the Competitiveness Council has made references to open coordination as a way forward. At the same time, there is a proliferation of objectives attached to the better regulation agenda. Looking at empirical evidence across Europe, the EU institutions have prioritized regulatory reform, but empirical analysis shows that at best there is clustered convergence across countries. Hence the risk is one of pursuing contradictory agendas camouflaged under 'common EU commitments' and 'Lisbon priorities'.

The final part of the paper is dedicated to a brief discussion of different scenarios for better regulation in relation to images of the regulatory state, and their implications for the legitimacy and the development of this policy area.
Pitfalls and limits for the coordination of social policies at EU level: the difficult relationship between evaluation and politics

Jean Claude Barbier

During the late 90s, and especially under the label of the "Lisbon strategy", a series of coordination initiatives were put in place by the EU, due to the Commission's active involvement (including certain effective political entrepreneurs). This gave rise to the invention and spreading of the "Open methods of coordination" (OMC), particularly the EES (European Employment Strategy) and the "MOC inclusion". In a second period, the momentum tended to subside, as the Commission's role became more modest and the 'inter-governmental' dimension rose again to importance. The second Kok report, the integration of the EES into a single list with the BEPG guidelines have recently concurred with the negative results of the referendums in France and the Netherlands to sideline the OMCs.

Evaluation in a way is at the core of this process, because in the period 1994-2005, when the EES was gaining form, the very basis of its effectiveness was its de-politicization. The choice for such and such a model of capitalism, as Radaelli has justly remarked, was not on the agenda. On the other hand, from the very early exercises of designing the guidelines, potential conflicts of values were present. These political and normative conflicts could not be solved through the use of evaluation or benchmarking. The referendums have shown that a process of politicization of the European social coordination is emerging. It seems that the Commission is in a way structurally unable to heed the messages from the referendum. In its latest communication to the Spring Council in March 2006, it presents the problem as if it were only a question of political communication and of disseminating "economic analysis". But "economic analysis" is precisely the stuff that has fed into the evaluation of the EES and other OMCs. One does not see why this sort of action (political communication about economic evaluation findings or ex-ante impacts) should really change the situation.
Current challenges for development evaluation

Dr Sean Conlin
Partner
ConMara Partnership, UK
Email: s.conlin1@btinternat.com

Panel Members:

Margaret Kakande
Head, Poverty Monitoring and Analysis Unit
Ministry of Finance, Planning and Economic Development, Uganda

Indran Naidoo
Chief Director, Governance Monitoring
Public Services Commission of South Africa, South Africa

Zenda Ofir
Board Member of AfrEA, South Africa

In recent years within the international development community, there has been international agreement on the Millennium Development Goals, enhanced co-operation in reducing poverty, and work to increase development effectiveness. The MDGs, agreed in 2000, have set new standards for multilateral organisations, donors and partner countries. For at least a decade, the international community has been developing partnership approaches to development assistance, such as sector approaches and the poverty reduction strategies. Some of these approaches, such as general budget support, have been the subject of intense debate within the international community regarding their effectiveness. Aid effectiveness has also been a subject of increased attention among donors over the last few years, with agreements to work towards better harmonization, alignment, and results. At the same time, policy makers have tried to reach beyond development assistance to consider the impact of other policy measures and private sector activities.

These changes in approaches in development policy have posed new challenges for evaluation specialists. In the early period of development in the 1960s and 1970s, assistance was delivered through projects and evaluation was focused on the efficiency and effectiveness of individual projects. Evaluation was usually a requirement of the donor that was providing the majority of project funds. The 1980s, with greater attention to structural adjustment policies, signalled the need for development evaluation to go beyond the project level. Since 2000, the task for evaluation has become even more complex. Evaluations need to deal with all kinds of policy-based approaches. Evaluation reports are required for all participating stakeholders and for purposes of lesson learning as well as for accountability. Results, increasingly, encompass policy influence and impact on the MDGs.

A panel will discuss the current challenges for development evaluation posed by these trends, especially since the agreement of the international community to work in partnership to achieve the MDGs. The panel will be composed of speakers who attended a Workshop at Sussex University preceding the conference, which was structured to enhance our common understanding of these challenges for development evaluation and to suggest ways forward.
Creating and developing Evaluation Organizations: Lessons learned from Africa, Asia, Americas, Australasia and Europe

Marco Segone
Regional Advisor, Monitoring and Evaluation
UNICEF Regional Office for Central and Eastern Europe (CEE) and the Commonwealth of Independent States (CIS), Switzerland
Email: msegone@unicef.org

IOCE (International Organization for Cooperation in Evaluation) decided to develop an 'organization package' consisting of resources for creating and developing evaluation organizations. The aim of this organization package is to help establish and/or develop successful evaluation organizations through the wide dissemination of case studies from all over the world on how such organizations can be created and managed.

14 case studies from Africa, Asia, Americas, Australasia and Europe are presented. Seven of the organisations are regional or national organisations from developed countries (including EES, AEA, CES); the other seven are from developed countries (including African Evaluation Association, Brazilian Evaluation Network and IPEN – covering the Former Soviet Union countries).

In each case study, the following issues are addressed:

- General information about the organisation, including organisational structure, number of members, and number and types of activities.
- Background on how the organisation started and how it developed.
- Three aspects or characteristics about the organisation that work best and three challenging aspects.
- Three major challenges met and their solutions.
- Three major lessons learned.
- Three key recommendations or tips on how to develop evaluation organisations of this kind.

Finally, a comparative analysis is presented, focusing on lessons learned extrapolated from the 14 case studies.

Marco Segone, in his function of IOCE Vice President (2003-2005) ensured the coordination of the initiative, and will present the major findings.
Towards a Human Rights Based Framework for Aid Evaluation

Marta Foresti
Overseas Development Institute, London, UK

David Booth
Overseas Development Institute, London, UK

Tammie O’Neil
Overseas Development Institute, London, UK

In this paper we analyse the synergies between human rights and the aid-effectiveness agenda set out in the ‘Paris Declaration’ in 2005. We do so by conducting an evaluative analysis of the Paris Declaration and its potential to achieve better and more effective aid, development results and objectives such as the MDGs. In particular, we consider to what extent human rights can contribute to the effective implementation of the Paris Declaration, by considering what are the key features of its theory of change and underpinning programme logic.

Defining the real potential, scope, role and impacts of the PD in delivering effective aid and contributing to development results will be the main task of monitoring and evaluating its implementation. In evaluation terms, the five commitments have a definite but largely implicit ‘programme logic’. In other words, there is a more or less shared informal theory about how the commitments, if fully implemented, would make aid more effective in contributing to development outcomes and the Millennium Development Goals (MDGs). Some parts of this theory are fairly well grounded in research or analysed experience – e.g. that country ownership of the policy agenda is a strong determinant of effectiveness. Others are more widely shared hunches that have not been tested and are often disputed – e.g. the belief that a results-based management is good for aid alignment and country ownership, or that aid harmonisation is invariably consistent with country ownership. In the paper we propose that the evaluation of the Paris Declaration presents a unique opportunity to consider the extent to which human rights can make a practical contribution in better defining the theory of changes of the Declaration and, in turn, its implementation. But how do we get there? Can human rights help to increase the likelihood that aid that is delivered and managed according to the Paris principles will actually result in sustainable changes in the lives of people?

We found that human rights have the potential to strengthen the implementation of the Paris Declaration in a number of ways:

- by better defining the scope and boundaries of the Mutual Accountability commitment, particularly in relation to strengthened domestic accountability in partners countries
- by placing equal emphasis on the processes leading to nationally owned policies, which must include all relevant institutions and sectors of society
- by using the international human rights framework as a starting point for Alignment and Harmonisation effort
- by ensuring that Management for Results is aimed at informing citizens as well as donors in a transparent and timely manner.

The paper adopts a very innovative approach by applying an operational and evaluative framework to analyse the potential synergies between human rights and development practice, overcoming some of the ideological and disciplinary barriers that often characterise this debate.
Ethical Aspects of Evaluation Practice or Ethics in Evaluation Practice

Dr Petri Virtanen
Senior Partner
Net Effect Ltd, Finland
Email: petri.virtanen@neteffect.fi

Associate Professor María Bustelo
Complutense University, Spain
Email: mbustelo@cps.ucm.es

Professor Thomas A. Schwandt
University of Illinois, USA
Email: tschwand@uiuc.edu

Professor Helen Simons
University of Southampton, UK
Email: h.simons@soton.ac.uk

Dr Jacques Toulemonde
Eureval, France
Email: toulemonde@eureval-c3e.fr

The topics of evaluation ethics, standards, and guidelines have received a great deal of attention in recent years in evaluation associations throughout the world. It seems that that while we can live in a world without standards and guidelines, we cannot live in a world without ethics. The papers presented in the symposium will explore the ethical dimension of evaluation practice, scrutinizing the ethical discourse promoted by evaluation theorists, the role and possible futures of evaluation standards (with or without an ethical dimension) as well as the current moral landscape of evaluation practice. The symposium will take up controversial aspects of evaluation standards as well as new competence demands in terms of ethics among professional evaluators. Interesting questions arise related to how to advance ethical practice in evaluation and how evaluators and future evaluators might be trained. Is there such a thing as a universal-theory of ethics in evaluation? Or is it only possible to prepare evaluators to act ethically in and through specific cases in practice? Other questions concern the role of standards in relation to ethics: is ethical practice too difficult to embody in codes and standards, or is it that codes and standards have a different function? These and other questions will be debated with a view to enhancing our understanding of how to advance ethical practice in the field of evaluation.

How can evaluators be trained in ethics?
Maria Bustelo

Norms, standards and guidelines have become very prominent in evaluation in recent years. Moreover, they have catalysed the debate around evaluation ethics. In addition to the role they can play in consolidating evaluation as a professional and academic field, they are supposed to guide evaluators in the ethical realm. However, their actual usefulness for this last purpose is under discussion, given different interpretations that might be made in the specific contexts in which they are applied. How can we then train evaluators and future evaluators in the ethical realm? Is there a universal-theory of ethics in evaluation that can be applied in all situations? Or it is only possible to
train in a context-specific "applied ethics" setting? In this paper I will address the challenge of ethical training for evaluators. I will argue that training for ethics needs to be situated in context but that there might still be a role for norms, standards and guidelines, as a reference for studying and discussing their different meanings and interpretations applied to different contexts and situations.

Maria Bustelo is Associate Professor of political science and public administration and Director of the Master on Evaluation of Public Programmes and Policies at the Complutense University at Madrid (Spain)

Can we live without ethics, guidelines, and standards in evaluation practice?
Thomas A. Schwandt

The topics of evaluation ethics, standards, and guidelines have received a great deal of attention in recent years in evaluation associations throughout the world. The purpose of this paper is to sort out the difference among matters of ethics, standards, and guidelines in evaluation while attempting to take into account the fact that these terms have different meaning in different languages. I will argue that while we can live in a world without standards and guidelines, we cannot live in a world without ethics.

Thomas A. Schwandt is Professor of Education at the University of Illinois, USA. He teaches courses in the philosophy of interpretive methodologies in the social sciences and theory of evaluation.

Ethical evaluation practice: principles, people and politics in action
Helen Simons

The rise of standards, guidelines and principles in professional societies is often coupled with an intention to advance ethical practice in evaluation. Yet Newman and Brown (1996) point out in relation to the Joint Committee Standards, that 'the word ethics is assiduously avoided' and Fraser (2005) has noted a similar lack of attention to ethics in the debate over standards in the Australasian Evaluation Society (AES). Is this because ethical practice is too difficult to embody in codes and standards? Is it that codes and standards have a different function? After briefly exploring the purposes principles and standards serve, this paper will examine what it means to be ethical in the policy and practice spaces in which we evaluate, drawing attention to the situatedness of ethical practice, the relational nature of ethical decision-making and the balance of political and professional judgement that is needed to act ethically in evaluation.

Helen Simons is Professor of Education and Evaluation at the University of Southampton (UK) where she specialises in programme, policy and institutional evaluation and the teaching of evaluation and qualitative research. She is currently President of the UKES and co-author of the UKES Guidelines for Good Practice in evaluation

Principles and practice: how do ethical principles contribute to delineate the evaluation profession?
Jacques Toulemonde

This paper builds upon the ten year old and still lively debate on standards and ethics within the French Evaluation Society (SFE). In a preamble to its seven principle charter, the SFE stresses that "Beyond the interests of the direct stakeholders, ... evaluations address the general public interest. Hence, they should abide by specific guiding principles". The most vividly debated principle is that of pluralism ("to take stock in a balanced manner of the whole range of legitimate points of view"). A first area of discussion is feasibility: is the pluralism principle workable at all? This point has been under discussion for three years, mainly on the basis of "ethical case studies" and lessons from experience. A second area of discussion involves the very definition of evaluation. Should the evaluation profession be delineated by referring to pluralism?

Jacques Toulemonde is partner of Eureval, a Lyon based private consultancy; co-director of the masters in public policy evaluation at Lyon Institute of Policy Studies, member of the International
Three enlightenments of evaluation ethics

Petri Virtanen

The Enlightenment with capitol “E” is something we often associate with the seventeenth and eighteenth centuries. This paper argues that there are three distinct enlightenments in terms of evaluation ethics. Enlightenments are revolutions in our thinking with reference to evaluation ethics. The first enlightenment described is associated with Michael Scriven and his contemporaneous thinkers of the 1970s and 1980s. The second enlightenment is incorporated with the evolution of various evaluation standards of the 1980s and 1990s. Finally, the third enlightenment refers to the idea of new conceptual definition of evaluation ethics – a definition emphasizing that the function of evaluation ethics is not to arrive at universal principles, but to contribute to the situational solution of practical problems in evaluation practice. Drawing from John Devey in particular, we might add that although we can today be somewhat guided by universal principles (such as evaluation standards) in the solution of at least some practical problems, few real problems can be solved by treating them as mere instances of a universal generalization. Good judgement and ethical behaviour are not entities that can be reduced to an algorithm. Contemporary moral landscape of evaluation ethics remain unclear as long as the conceptual boundaries and contents of evaluation ethics remain fuzzy. This is the task of third enlightenment to accomplish.

Petri Virtanen is Senior Partner and chairman of the board at Net Effect Ltd., a Helsinki-based private consultancy, and he is involved in evaluation and development projects commissioned by the Finnish ministries, the European Commission and the World Bank. He was President of the Finnish Evaluation Society during 1999-2003.
The five data box conceptualization: a comprehensive evaluation framework

Professor Werner W Wittmann
University of Mannheim, Germany
Email: wittmann@tnt.psychologie.uni-mannheim.de

The five data box conceptualization synthesizes different schools and theories of evaluation research, e.g. the Northwestern school's emphasis on reforms as experiments and quasi-experiments, Cronbach et al. focus on correlational research and generalizability, the importance of stakeholder perspectives and symmetry principles between what is treated/taught and tested. The evaluation box encompasses different stakeholder perspectives, which have to be implemented and assessed in the criterion box. The experimental and the non-experimental treatment boxes map the intervention/program dimensions as potential causes for the outcomes in the criterion box. The predictor box assesses the status before any interventions. The boxes are ordered according to the time arrow and thus allow path-analysis to test causal relationships. The whole framework has been successfully applied to several programs and interventions. Examples from health/rehabilitation programs in Germany are presented as well as a meta-analysis of 65 studies, which more or less applied that framework.

References:


Evaluating a disease management service in the UK: governance, challenges and standards

Dr Pejman Azarmina
Outcomes and Evaluation Lead
Pfizer Limited, UK
Email: Pejman.azarmina@pfizer.com

This paper describes the challenges of evaluating an innovative disease management (DM) service in the UK. Disease management is a coordinated system of care that is founded on the pillars of ‘the chronic care model’ and includes patient empowerment and supported self care. Ideally, users of these services are more informed about their conditions and take action to improve their healthy lifestyles, take their medications correctly and get recommended tests and services. They also know how and when to call their doctor and follow-up with their care plans and appointments. This service is provided by ‘Care Managers’ who are usually registered nurses that are based in a call centre and are responsible for the care management of an invited population. The goal of disease management services in the UK is to improve health outcomes for patients with long-term conditions and to lessen health inequalities across the population.

The evaluation lessons learned from randomised DM trials and a bespoke DM service will be used as illustrations and the challenges of having a comparator group will be presented. Additionally, the need for development of a UK-specific evaluation plan is discussed in the light of the NHS research governance and clinical governance strategies and the UK’s National Programme for IT. This paper would finally propose a standard by which a disease management service could be evaluated against. A proposed standard evaluation plan will be presented which can produce a hologram of evidence using a wide range of qualitative and quantitative approaches (e.g. clinical audit, health equity audit, benchmarking, focus groups, satisfaction surveys and management report). The required data is obtained either from clinical information systems or by conducting satellite research projects out of partnerships between the NHS, private sector and academia.
Abstract 021

Designing and Implementing an Evaluation System for the Canadian Military Family Service Program

Dr Peter Gabor
Faculty of Social Work, University of Calgary, Canada
Email: gabor@uleth.ca

Alla Ivask
Senior Manager, Program and Policy Development
Directorate of Military Family Services, Canada

Sue McCormack
Field Operations Manager
Directorate of Military Family Services, Canada

In 2004, the Canadian Directorate Military Family Services, the corporate manager of the Canadian Military Family Services Program, launched an initiative that reconceptualized how the program was to be delivered. The delivery system consists of 42 Military Family Resource Centres, located throughout Canada and in selected locations in Europe and the United States; the centres are located in a wide range of community contexts and vary greatly in service volume and budget.

An integral part of the initiative was to develop an accountability system with two main purposes: (1) ensure high quality, accountable programming, congruent with the principles of family support, and (2) to provide for ongoing program development. To implement the accountability framework, an evaluation system with capability to capture key data within centres and to return relevant information to each centre was sought. At the same time, the capacity to aggregate results across the entire system was important to provide an overall picture to the directorate.

The major evaluation challenge was to develop a single system that could be implemented in organizations of varying capacities, from large, multi-funded centres serving thousands of families to small organizations serving a few dozen. The system had to be manageable yet produce information that would be relevant both to individual centres and to the corporate managers.

An evaluation system was designed over a two year period, using a consultative approach that included representatives from the centres, the directorate, and an evaluation consultant.

This paper will describe the evaluation needs of the system from the corporate perspective, outline the evaluation challenges posed, discuss the evaluation solutions and the implementation of the evaluation system.

This paper speaks to the challenges and solutions involved in the evaluation of a larger system (Stream 5). Further, it highlights a number of critical connections; most importantly it addresses the connections between evaluation, accountability and program development and between local level and higher corporate level evaluation activities.
The Use of Ethnography in the Evaluation of Scottish Healthy Living Centres: Enacting and Managing Everyday Evaluation Practice

David Rankin
Research Fellow
Research Unit in Health, Behaviour and Change
University of Edinburgh
Email: a.d.rankin@ed.ac.uk

A two-phase process evaluation of a sample of Scottish Healthy Living Centres (HLCs) has been underway since 2002. Six case study sites were purposively selected to take account of different contexts (e.g. focus, geography and structure) that are found in the wider programme. Using an ethnographic approach, phase one explored pathways between activities, processes, contexts and outcomes, while phase two concentrates on HLC attempts to address inequalities and attain longer-term sustainability. I examine how an ethnographic approach enhances understanding of the complex and dynamic processes occurring within HLCs; and how such an approach affects participants' reactions to my role as an evaluator and, in turn, how these reactions impact on the practice of evaluation.

To begin, I provide an overview of the HLC programme and the rationale for adopting an ethnographic approach to evaluation. This is followed by an explication of the methods used, including: in-depth interviewing; participant observation; documentary review; and, over time, an increasing number of researcher/participant interactions. These 'ethnographic encounters' include: informal interactions (e.g. telephone conversations to plan meetings); casual office interactions; participation in wider community events; and impromptu meetings (e.g. at conferences). Using fieldwork examples, I illustrate how ethnographic approaches have contributed to the development of evaluator/participant relationships and how these facilitate understanding of the cultural complexity of HLC processes in social contexts.

Second, I examine how choices of methods, personal characteristics of the evaluator and practical constraints of specific contexts influence evaluation practice. Everyday situations, placed in context, are used to illustrate how participants view the evaluator role. I then examine my responses to participants' construal of the evaluator role in attempts to ensure smooth running and continuation of the evaluation over time. More broadly, this study demonstrates the evaluator's need to be continually reflexive and responsive to the everyday enactment of doing an evaluation.
From drainpipes to integrated systems. Evaluations for policy efficiency and democratic development in different policy sectors

Dr Åsa Sohlman
Sohlman Senior Consultants, Sweden
Email: sasohlman@privat.utfors.se

In collaboration with:
Marja Lemne
Senior Advisor
The Swedish Agency for Public Management, Sweden
Staffan Sohlman
Sohlman Senior Consultants, Sweden

The Swedish system of evaluation at Government level is developing from a traditional ad hoc system based on Government Commissions of inquiry towards a system of management by objectives and results. The managing of the abundant Swedish government agencies is combined with feedback reporting and evaluations from these agencies.

In this paper we investigate how the new and the old system fit in with the development in different policy sectors in Sweden. The sectors studied are i.a. labour market policy, educational policy, defence policy and policies for the integration of immigrants. Some of these policy sectors work in a narrow, hierarchical way while others are more fragmented. There are examples of integrating elements in the fragmented policy sectors but none of them are fully integrated nor do they all rely on evaluations.

Sector structures and accompanying evaluations systems are described and solutions to policy challenges discussed.

As for the interface between evaluation systems and policy, evaluations can be discussed on the one hand for small, continuous policy changes on the other hand for new policy reforms. In the first case one trend is towards evaluations commissioned by the government and supposedly used for policy adaptations. This approach may be relevant for the drainpipes. For complex crossover sectors new methods are being tried out, e.g. new methods to commission and handle evaluations.

Policy reforms are not only inspired by evaluations. If evaluations are to play a role for new policies, which seems legitimate for policy efficiency as well as for democratic development, it is reasonable that the official evaluation systems are embedded in larger systems of dynamic research and external and internal reporting. Arenas for debate and discussion of interpretation of evaluation results are also desirable.
Adapting Evaluation to Changing Political, National and Individual Contexts

Frances Salt
National Manager Reporting Services
Education Review Office, New Zealand
Email: frances.salt@ero.govt.nz

Isabell Sinclair Irwin
Education Review Office, New Zealand

The Education Review Office (ERO) was established in 1989 at the time of the implementation of significant education reforms in New Zealand. It is responsible for reviewing and reporting publicly on the performance of schools and early childhood services. Over the last 16 years ERO has continually adapted its evaluation methodology in response to changes of government, changes of education policy, changes in the school and early childhood sector, and management changes within the Office itself.

This presentation explores the rationale for each evaluation approach, the advantages and limitations of each, and the reasons for consequent changes and adaptations. It considers possible directional changes in the short to medium term and why these might be desirable or necessary.
Promoting the utilization of effectiveness evidences and guide lines in the sector of health prevention

Professor Liliana Leone

Director
Studio CEVAS (Consultancy and Evaluation in the Social sector), Italy
Email: leone@cevas.it

The paper will focus on the utilization at different levels (national, regional and local) of evaluation evidences in the sectors of health promotion and licit and illicit drug prevention. The presentation intends to address the following issues:

- Different stakeholder and utilization of evidence from evaluation: limits and strengths of systematic reviews;
- The suggestion of "realist synthesis": a more useful way to synthesize the evidences of effectiveness;
- The incorporation of evidence inside a new evaluation;
- Reinforcing the policy making processes and the planning of prevention on the base of evaluation evidence;
- The refining and diffusion of recommendations drawn from the evaluative research and systematic reviews to practitioners and policy makers.

The above issues will be considered and dealt with thanks to the presentation of a case study evaluation characterized by strong participatory processes. Different stakeholders were involved in a project over a 3 year period (years 2003-2005): the national department and regional administrations of north Italy (Lombardia, Liguria, Piemonte), the management of 11 Local Health Units (LHU), the health and social workers of departments for drug addiction including those of non profit organizations, boroughs, schools, local agencies involved in youth policies and international agencies.

The initial step of the evaluation was identifying and classifying evidences of effectiveness drawn up from the most important existent meta evaluations and guide lines. Although the systematic reviews had in great part ignored the role of theory in the explanatory processes, we tried to classify the results through the concept of "strategy of prevention" that has been in part assimilated to the concept of the "family of mechanisms" (Pawson 2006). The concept of "strategy" cuts in a transverse way programmes implemented in different fields of prevention policies (i.e., road safety, drug and smoke and alcohol addiction, health education.); for this reason was easier to communicate the results and recommendations of evaluation and strengthen the collaboration between sectors and departments.
Left holding the baby

Dr Dione Hills
Principal Researcher Consultant
The Tavistock Institute, UK

Co author
Emma Wilkinson
Researcher Consultant and
The Tavistock Institute, UK

This presentation looks at the role of the evaluator in those programme evaluations where the programme continues over a period of years, but the policies that formed the background have changed and new policies have become a priority. The evaluator can be left holding the ageing baby as its supporters go on to new creations.

The presentation will make use of the healthy living centre programme, launched in 1998 with a budget of £300 million from the New Opportunities Fund (now the Big Lottery Fund). Funding 350 projects across the UK, it represented a major investment in the promotion of health and wellbeing at a local level and funded projects for up to five years. The healthy living centres were heralded as flagships of a new strategy to address health inequalities at a neighbourhood level.

Changes in the wider policy agenda and local context since the programme was set up have had important implications for local projects and the potential for their sustainability.

As well as making changes to the evaluation strategy in order to ensure that lessons captured from the programme had relevance to the new policy agendas, the evaluators also found themselves dealing with requests for advocacy, reminding stakeholders about the original role of the programme and drawing the attention of policy makers to the possible role of these centres in consultations on new strategies and policies.

We will discuss this and other examples in the context of knowledge utilisation and methodologies which promote the capture and dissemination of evaluation findings. The ethos of stakeholder engagement and participation and its relevance to the policy research field and the implications for the evaluator skill set will also be explored.
A diagnostic about Evaluation use in France

Bernard Perret
Conseil général des ponts et chaussées
Ministère de l'équipement, France
Email: Bernard.perret@equipement.gouv.fr

In France, the institutionalization of evaluation by the end of the years 1980 was supported by a discourse on “Democratic Evaluation”. Evaluation was supposed to bring more transparency and rationality into the public debate. Fifteen years after, it can be seen that this expectation was too optimistic. Evaluation is practically never mobilized as a specific input in policies related controversies. It has not been used either as a current budgetary and/or accountability tool. The fact that the new budgetary law – instituting programs based budgeting and performance indicators - gives no explicit role to evaluation, is illustrative of this situation. In fact, elected people, medias and public opinion see no difference between an evaluation and any other source of information or expertise – academic study or research, administrative or parliamentary report, individual statement by a supposed specialist, etc.

Evaluation, however, has developed inside public administrations. So the question is: for which kind of uses and effects can it be credited? Observations made in various public administrations suggest that the most convincing uses of institutionalized evaluation are managerial, but not in the narrow sense of an efficiency tool. Through evaluation has been invented, at least in some places, a new kind of dialog among policies actors and stakeholders about public action issues. The pluralistic steering committees (Instances d'évaluation) play a crucial role in this perspective: they give place to discussions about policies objectives and strategies mediated by discussions about methodological and informational issues. These kind of evaluations may help to take administrative decisions that have more chance to be appropriated by relevant actors and, thus, have more chance to be effective. The problem with this use of evaluation is that it has no visibility outside the policies actors sphere.

The aim of the presentation is to discuss and illustrate this diagnostic through various examples.
Interactions, instruments and aims in VFM audit: a cross-national study

Stuart Astill
Research Fellow
National Audit Office, UK
Email: s.j.astill@lse.ac.uk

Edward Page
London School of Economics, UK

In order to be able work towards improving practice we need to understand that the systems within which we work are constrained by their external context and, quite simply, by the work we do. Other elements of the system however are, in effect, choices. With this knowledge we can concentrate efforts towards change in arenas that are amenable to change.

The paper summarises research examining the practice and context of the UK National Audit Office and, as a comparator, the French Cour des Comptes in order to describe the complex and interlocking system that has evolved over time to address the conflicts inherent in prioritising and achieving aims, choosing instruments, establishing and managing interactions.

By comparing and contrasting the two institutions the goal is to see in which arenas change could be most profitably pursued, where opportunities exist for exchanging best practice and how great is the potential for working together.

The different roles and approaches of the two audit institutions and the individual contexts of the two countries, as well as the frame of research, mean that the study does not seek to judge the institutions but to untangle the framework of interactions and discover how the elements of the VFM audit system fit together.

The research aims to create an increased understanding of how future choices have an impact across the system that comprises the VFM audit function.
The Finnish R & D Evaluation: is there a real accountability perspective?

Dr Timo Oksanen
Senior Auditor
State Audit Office, Finland
Email: timo.oksanen@vtv.fi

R & D expenditure in Finland has increased at annual rate of 13.5 per cent while average annual real grow in all the EU countries has been 3.4 per cent. Besides that Finnish R & D system has encountered many changes and challenges since the mid-1990s.

Though the evaluation and research of the R & D is growing by and by, our knowledge of the impact of the R & D is still rather narrow. Meanwhile “evaluation enterprise” at the whole in Finland – the amount of evaluation firms, researchers, consults, activities and reports – is rapidly growing. In this situation the Finnish Ministry of Finance started 1997 an independent expert analysis "Evaluation Support of the Finnish Central Government". The implementing of the recommendations of analyses has not been finished until today (except e.g. the annual accountability reporting system of the government to parliament, which has been reformed).

One of the main strengths of the Finnish R & D system has been the multisectoral innovation policy which has been developed from the beginning of the 1990s. It seems that the evaluation of the Finnish R & D cannot produce a systematic knowledge-base to the innovation policy (see OECD, so called Monet project). However, an up-dated analyses in which the problem is studied as a part of more focused problem of the governance of the "accountability perspective to evaluation" (Chelimsky) of R & D is still mostly lacking.

The State Audit Office started at 2005 a pilot audit in which some important aspects of the systemic problem mentioned afore will be met. The focus of the audit has been put on to next three questions:

1. By which means, instruments and practices, official or nonofficial, the evaluation of the Finnish R & D has been linked in to policy arenas and practices?
2. By which means the rational utilizing of the R & D evaluation has otherwise been ensured?
3. How the evaluations of the Finnish R & D programs has in some case’s (e.g. in the cases of some Finnish national cluster programs) been utilized?

Our audit processes started at the beginning of the 2006. We think that our data collecting processes and preliminary analyses are finished until Conference (4-6 October 2006) so that we can describe at least some important points of the questions 1 and 2.

Dr Prof Maurice Baslé
Faculty of Economics
University of Rennes 1, France
Email: maurice.basle@univ-rennes1.fr

The paper is a survey of 160 French evaluations reports collected by the national agency DATAR.

Some lessons are presented and recommendations for the next generation of French contracts (2007-2013).

The crucial weaknesses of data bases and knowledge bases are also presented. More theoretical, there is a crucial lack of theory of action, diagram of impact must be developed for the future.
Evaluating the nature of public policy making: the scope for synthesis

Dr Paul Burton
Senior Lecturer and Head of School
School for Policy Studies
University of Bristol
Email: paul.burton@bristol.ac.uk

This paper considers the ways in which evaluation is used in the process of policy making and takes this as a starting point for a critical review of the ways in which the process of policy making is conceptualised. Alternative conceptions of this process are compared and the scope for synthesis is assessed.

The paper concludes by proposing a new conceptualisation of policy making and setting out some preliminary thoughts on the implications of this for the theory and practice of evaluation.
Evaluations of School Performance: Meeting the Needs of Various Audiences

Frances Salt  
National Manager Reporting Services  
Education Review Office, New Zealand  
Email: frances.salt@ero.govt.nz

Isabell Sinclair Irwin  
Education Review Office, New Zealand

The New Zealand Education Review Office (ERO) is responsible for evaluating the quality of education provided for students in New Zealand schools and reporting these findings to individual schools, to parents and communities, and to the New Zealand Government. Although it is a commonly held premise that evaluations cannot effectively serve several purposes this presentation will discuss an innovative evaluation methodology implemented by ERO for its reviews of schools. This methodology has the flexibility to provide useful evaluation findings to all three audiences.

The presentation will describe that ways in which the evaluations are undertaken to provide evaluative information for various audiences, how this information is promulgated and what use audiences make of it. While there is considerable evidence that schools and education policy makers use the information, the challenge for ERO now is to make information more readily available to parents and school communities.
Evaluation as policy evolves

Richard Gaunt
CRG Research Ltd
Email: richard@crgresearch.co.uk

Evaluation of well-established projects and programmes is vulnerable to the criticism of being ‘too late to make a difference’. Evaluation at too early a stage is vulnerable to the criticism that ‘not much has happened’.

The paper argues that evaluators can play a valuable role as policy evolves, but need to be aware of their limitations. In principle, evaluators can play important roles as policy evolves through clarifying issues like ‘what would success look like?’, ‘what are early stage indications of productive avenues to follow?’ Various challenges arise, however. By giving views about what may be productive avenues to follow, evaluators may start to tread on areas like legality, propriety and organisational feasibility. They will also be operating well out of usual comfort areas where there are robust sample sizes to determine, counterfactuals to assess and carefully structured research questions to address.

These issues will be developed through practical experience, in particular, of providing evaluation support for initiatives to help long term economically active people gain productive employment in the construction sector.

Although formative evaluation issues needed to be addressed before the end of the project, a wide range of early stage challenges had to be faced including contractual parameters, the likely reaction of contractors and other stakeholders, and the ‘true’ picture lying behind official labour market statistics.
The role of side effects in public policy making: environmentally harmful subsidies

Kris Bachus
Researcher
University of Leuven, Belgium
Email: kris.bachus@hiva.kuleuven.be

Side effects are policy effects that are not intended by the policy. They can be either positive or negative, and they can take place either in the studied policy field or outside of it. This paper will look at the issue of side effects from two perspectives.

Firstly, the paper gives an overview of a side effect evaluation study we carried out on the topic of ‘environmentally harmful subsidies’: for this research project, we focused on the environmental side effects of economic subsidies, both in the agricultural and the transport sector.

Secondly, we will supplement the results of this evaluation study with some generalising considerations on the importance of side effects in the making of public policy. These considerations will be based on the subsidy study and on a number of other studies we carried out on the topic of side effects of environmental policy.

It appears that, although there is a broad consensus about the importance of environmental policy in general, environmental policy measures have always been very controversial. We claim that this has to do with the (presumed) existence of negative side effects; lower competitiveness of industry and negative employment effects are the most heard fears.
The reflective practitioner?

Jennie Fleming
Principal Lecturer (Research) and Director of Centre for Social Action
Centre for Social Action
De Montfort University, UK
Email: jfleming@dmu.ac.uk

Co author
Thilo Boeck
Senior Research Fellow
De Montfort University, UK

Our paper draws on two specific experiences – the SARP project a multi-agency project funded by the HDA (Fleming and Boeck 2005) and the evaluation of a local Children’s Fund. In these projects we have found that the ability of practitioners to engage in the process of critical reflection and evaluative practise is very variable. We would agree with Coote A et al who write ‘the value of involving practitioners in evaluation and learning from their experience should be more widely recognised – and skills and techniques must be developed to enable this to happen.’ (pxiii), however this has proven to be quite a challenge in some of our work.

By enabling practitioners to develop their own methods for evaluating their work we found that it is crucial to support the projects to be flexible and evaluation to be tailored to the needs of each project. However we have also found a number of tensions in this approach:

- evaluation is often an afterthought and so reflection is limited as evaluation is reactive not proactive
- frequently evaluation and satisfaction surveys are seen as the same thing by practitioners
- frameworks for evaluation and the development of indicators (e.g. EveryChild Matters outcomes) are imposed centrally, this does not this match up with a participative agenda.

Since evaluation should inform and improve practice and be a catalyst for learning this can be a problem for organisations both in developing a stronger learning culture and also in them being accountable for the public monies they are receiving for their organisation. Despite Much good practice the variable ability to be reflective about this can influence their ability to attract further funding and develop their practice according to national demands. Our paper will explore some of these tensions and present some ways towards addressing them, whilst recognising we are a long way from solving them all.

How evaluation can contribute to improve practice in the professions. The case of a Danish ICT programme

Thomas Bjornkilde
Managing Director
Ramboll Management, Sweden
Email: Thomas.bjornkilde@r-m.com

Nicolaj Ejler
Ramboll Management, Sweden

In 2001, The Danish Ministry of Education initiated the programme ITMF (IT and media in the primary school) with a budget of 50 MEUR to improve the use of ICT in primary schools in Denmark. The main part of the budget was allocated to approx. 90 pilot projects. From the beginning of the project, an external evaluator was attached to the programme being responsible for supporting the implementation of the projects and facilitating exchange of knowledge between the projects. A key instrument for this purpose was an internet based self-evaluation system. The system was designed on a “theory of change” approach and provided the project holders with tools and guidelines for various parts of the project cycle, e.g.:

1. Initiation phase: Developing theory of change, establish a log frame and an evaluation plan
2. Implementation: Monitoring and process evaluation
3. Finalisation: Final evaluation

The main target group for the self-evaluation system was teacher educated project managers. In order to develop their evaluation skills, two one-day workshops introducing evaluation in general and the self-evaluation system in particular were held.

An evaluation of the utilisation and impact of the self-evaluation system showed mixed results. 30% of the project managers have used the tool to a high degree and assessed that the tool was supportive for their project management. But the remaining 70% did not use the system. First, it was not mandatory to use the system. Second, the project manager role was more time consuming than expected. Final, self-evaluation is still a new phenomenon in the Danish educational system at primary level.

The experience from the programme demonstrates that self-evaluation tools have potential to improve practice within a profession. But, in order to do so more attention needs to be paid to evaluation capacity building and programme management. Clearly, there is a need for professionals to increase their evaluation skills. But it is equally important that programme managers create incentives and provide the resources needed for developing the evaluation capacity.
Steps towards Relational Evaluation of Organizational Development

Katri Vataja
Researcher
National Research and Development Centre for Welfare and Health (STAKES)
Email: katri.vataja@stakes.fi

Juha Koivisto
Researcher
National Research and Development Centre for Welfare and Health (STAKES)

Co author
Riitta Seppänen-Järvelä

The paper presents a relational approach on the evaluation of effectiveness of organizational development activities. The approach bases on relational ontology and aims thus to cross the firmly established ontological diversities within the social sciences, such as objective versus subjective, actor versus structure, intervention versus context, process versus outcome, inside versus outside and summative versus formative.

The paper analyzes development activities as complicated and mediated local activities and interactions between different human actors. Organizations and work communities are studied as sociomaterial networks and relations which are continuously constituted by the human and other elements.

The paper argues that the continuous production of social structures and the development activities that attempt to change the practices and structures can be empirically analysed by combining different qualitative and quantitative research methods. In practice, to evaluate the effectiveness of the development activities, the organization has to be followed before the development activities, during the development activities and after the development activities. The research methods are not decided a priori, but they are rather chosen individually in every case. The relational approach is illustrated by a case study where the development activities in twelve Finnish social welfare offices are evaluated and where the evaluation is used to facilitate the development.
What can constructionist evaluation tell us about the learning individual and organization?

Robert Dalziel
Research Fellow
Institute of Local Government Studies
University of Birmingham
Email: r.dalziel@bham.ac.uk

This paper focuses on the different connections between programme evaluation, how people view the world around them, and the social construction of meaning. The aim is to show how the constructionist perspective, which underpinned the evaluation of a local authority senior management development programme, provided novel insights that improved stakeholder understanding of the operation and value of the programme. The paper draws on the literature on evaluation to examine critically the assumption that individual development will result in organizational development.

The paper will also draw on the literature on social policy evaluation and the social construction of meaning to show how, paradoxically, the ‘modernizing agenda’ and the capacity building in organizations that it promotes has failed to effectively value difference and stifled innovation. There has been a concentration of effort on the creation of organizational structures to deliver training and the evaluation of organizational performance at the expense of understanding the interconnectedness of context, culture, values and evaluation.

The paper will argue that a constructionist evaluation encourages a reflexive process that carefully considers how different political tensions and conflict over what learning is appropriate or acceptable affect stakeholder perceptions, feelings and emotions.
Evaluator or Consultant: Constructing a role from different identities in evaluating a single organisation

Frances Abraham
Principal Consultant and Senior Researcher
The Tavistock Institute, UK
Email: f.abraham@tavinstitute.org

From undertaking organisational reviews for a number of organisations, it seems that evaluating single organisations requires a broad set of organisational theories drawn from organisational behaviour and consultancy competence as well as theory and competence in evaluation. In this particular case, the organisation wanted both an evaluation of its services and a review of the organisation’s capacity to deliver its services in a sustainable way.

The evaluation of services provided by this organisation included data collection and feedback from a broad set of stakeholders, including regional sponsors of organisational streams of work, service users or beneficiaries and a range of intermediaries, as well as extensive documentary review. All methods and perspectives reported high levels of performance by the organisation.

However, despite limited access to the organisation being offered, a range of difficult strategic issues emerged in relation to the organisational review. These included the imminent departure of the chief executive, a high degree of contest between different staff groups around the espoused organisational mission and threats to future funding. Each of these issues had become unmentionable in the organisation to the degree that what could not be talked about was clearly an additional issue in this organisation, as the different interests closed down discussion of each strategic issue in turn.

Making sense of what was unmentionable and why it had become so required discourse analysis at the organisational level and organisational political analysis drawing on frameworks provided by Mintzberg. Introducing concepts and frameworks which might enable the organisation to break free of its habits of silence and covert contest required dialogic interventions and political action from consultancy competency frameworks in a role combining both evaluator and consultant identities.
From whence did we come? The emerging identity of the European Evaluation Society

Elise Ahn
Doctoral student
University of Illinois, Urbana-Champaign, USA
Email: eliseahn@uiuc.edu

After being founded in December 1994, the overarching goal of the European Evaluation Society (EES) is to “promote theory, practice and utilization of high quality evaluation” in (but not limited to) the European area. In light of the dominant outcomes and management-oriented evaluation paradigms by European Union (EU) funded projects, does the EES provide a space where other types of evaluation (or evaluation orientations) have space to grow and/or develop? Moreover, where is (or should/could) the intersection of theory and practice? And what is the effect of theory driven evaluations at different levels in the reporting hierarchy?

This paper looks at the programs for the past biennial EES meetings to see any patterns pointing to how the EES is emerging as a professional evaluation organization and as such a space. By looking at the session topics and individual paper break outs, it will be interesting to see 1) the development of the EES in terms of conference content/emphasis and 2) whether the EES could or is negotiating the role of evaluation in light of this dominant evaluation paradigm utilized by the EU.

The programs are observed from five different angles, mainly using descriptive statistics: geo-political locale (EU/Non-EU, European/Non-European, National-Government/Non-Governmental), middle theory topics (democracy, accountability, development, and governance), practice related (tools, practice, instruments, and standards), and lastly, language (English/Non-English).

By looking at the content of the conferences and whether this is changing over time, this paper hopes to offer a few observations on how the EU is shaping the discourse in the European evaluation field.
E-valuation and European Value: On the Values of Evaluations

Andreas Kolodziejak
Evaluation Manager
European Commission
Email: andreas.kolodziejak@ec.europa.eu

Values play by definition a large role in evaluations although many would argue that evaluations take place using the special policy objectives specified in the applicable regulation.

Evaluations in the European project take always place in a societal framework, but at the same time they also take place within a more limited programme and policy framework. Concerning the latter it must be noted that the frequent change of policies bring their characteristics close to projects. The two frameworks imply a different meaning for European values in the evaluations. This is important as the European project has included general policy principles like sustainability (Gothenburg Council).

The classical European evaluation approach based on specific programme or policy objectives as evaluation criteria, for reasons set out in the paper, over the years becomes more and more problematic to apply. An example is the case where the special policy objectives change over time or where European values are needed for their interpretation to make these objectives work as criteria in practice. The problems are not only in producing the evaluation but especially in its dissemination and communication with the European public and politics.

The paper investigates how evaluations in the European project would look like if European values rather than programme and policy specific evaluation criteria would be used. This would have a number of advantages like a shorter preparation time and clearer and deeper communication with the public. The latter is a key to the need for using general European values in the evaluation process. If evaluations are more like internal management tools the need to include e-values is much smaller.

The paper finally deals with the subject of the aim of evaluations in the European project and with which adaptations that would have to be made to fulfil the societal mission of evaluations.
Efficiency: a challenging question for evaluators, even more at European level.

Svend Jakobsen
European Commission
DG Budget

Jacques Toulemonde
Eureval

Lars Høberg
Rambøll Management

When the European Commission’s evaluation policy was launched in the mid 90s, the initiative came from DG Budget and the focus was very much on expenditures. One of the basic documents was entitled “Spending more wisely”. Meanwhile budgetary use of evaluation findings has been institutionalised as part of the Commission’s planning and budgetary cycle and the Council and the Parliament welcome evaluative information as an input to budgetary decision making.

Despite this high level of demand, estimation of efficiency remains a difficulty in EC evaluations, which often deal with complex interventions. This perception has been confirmed by a recent study of the use of Cost-Effectiveness Analysis (CEA) in the about 600 evaluation reports finalised between 2000 and 2005.

CEA is the typical approach for assessing efficiency. It allows evaluators to assess whether the cost of achieving a desired effect is reasonable, if compared to an appropriate benchmark. The comparison may apply across different components of the intervention, with an outside good practice, or with a hypothetic alternative strategy of implementation.

In practice, the canonical use of CEA is rather limited in the real world of evaluation and this applies to evaluations at European level as well. Evaluators tend to rely on CEA when interventions are simple, i.e. when a few cause-and-effect chains are connected to one major effect (e.g. jobs created, CO2 emission limited, etc), and when standard indicators are available.

The feasibility of CEA must be seriously questioned in the case of complex interventions, with co-funding arrangements, management through partnerships, multiple activities, and/or multiple intended effects. In such instances, which are frequent in the EC context, evaluators need to develop alternative methods, and in fact they have explored a range of pragmatic solutions.

The presentation is a contribution to describing the state of the art and the promising options in the challenging domain of evaluating the efficiency of complex interventions.
No abstract available
Knowledge logic and action logic: the construction of an evaluation tool

Johan Stessens
Higher Institute for Labour Studies - Katholieke Universiteit Leuven, Belgium

Gerlinde Doyen
Higher Institute for Labour Studies - Katholieke Universiteit Leuven, Belgium

Professor Patrick Develtere
Head of Research Group
Higher Institute for Labour Studies - Katholieke Universiteit Leuven, Belgium
Email: patrick.develtere@hiva.kuleuven.be

Co authors
Bénédicte Fonteneau
Ignace Pollet

All working for the Higher Institute for Labour Studies (HIVA, Catholic University of Leuven, Belgium Belgian NGOs are increasingly obliged to demonstrate the achievement of their objectives and are confronted with multiple internal and external challenges. Hence, attention paid to evaluation increased drastically over the last decade. Simultaneously with the evaluation process, several Belgian NGOs also initiated a general reflectivity process considered indispensable and vital for the organisations' survival and progress. In Belgium, these processes were facilitated with the obligation of spending 1% of the budget they receive from Belgian government to evaluation. In order to combine both processes some Belgian NGOs consulted the scientific community. In this way a confrontation between the knowledge logic of academia with the action logic applied in the NGOs came about.

In this paper, we try to clarify the principles and approach used by the authors’ research team in this joint NGO-academia undertaking. Starting from the logic of the actors and their perceptions on the evolution of their capacities and activities, the authors combine the rather essentialist approach with an immanent, scientifically underpinned, approach. The incorporation of two measurement periods (T0, T1) makes the approach multi-annual and dynamic since it allows for a confrontation of the stakeholders' perceptions (object of the evaluation or external) and the set of indicators.

We develop the advantages and limits of this approach while answering the following inevitable questions: (a) how to measure results and impact?; (b) how can NGOs be both commissioner and object of the evaluation?; (c) who is to participate in this exercise? And, (d) how to combine a factual analysis with the interpretations of the actors.
Evaluation of relationships supporting migrant activities

Professor Enrico Tezza
Senior Research Manager
ILO Turin Centre, Italy
Email: e.tezza@itcilo.org

A study by the Evaluation Unit of the ILO Turin Centre.

The study deals with the social relationship as an inclusion strategy. Since social network analysis (SNA) measures the social relationship, the use of SNA could be a key skill in tackling the inclusion of migrant workers. The evaluation is of a project run by the International Organization for Migration (in 2003 and 2004) to provide training to 2000 migrant workers as part of public placement action.

The social inclusion of migrant workers depends on the social network they inhabit. The effectiveness of local migration policies depends on the social network as well. Since the network is important for both the demand and the supply of "social inclusion", the understanding of its structure and functioning is crucial for designing coherent local migration policies. The emphasis on relations derives from the principle that social action is "inter-action" among social actors. According to Georg Simmel, social relations are central to how society works. Relations are the gateway to social ties, contacts, "rela(c)tion" with others. These "relations" become the object of specific study.

After presenting the network analysis structure (purpose, method, techniques and tools, respondents and coverage), the report outlines results on nodes, graphs and matrices (case by case, and affiliation by affiliation) processed using Ucinet software, together with other statistics obtained by SPSS, indicators of centrality and of structural balance. A network expansion coefficient and a network reinforcement coefficient are discussed as well.

Results confirm the premises behind the study. All the training agencies used a migrant target group support network for recruitment and placement, albeit to different extents. The central role in the recruitment of migrant workers was played by organisations in the voluntary sector. The relations listed for job placement reveal a network configuration characterised by agents active in job-seeking. Firms and organisations that are responsible for introducing migrants to the world of work feature prominently. There is a distinct predominance of private companies. Trade unions for migrants were contacted frequently for recruitment, but not often for occupational inclusion.

Migrant policies need good relationships: among the institutional agencies that design and support the policies; among the regional agencies that implement them; between institutional agencies and regional agencies; between agencies and migrants; among migrants themselves; and between migrants and the host population. Network analysis can be used to investigate these relationships and the communicative processes involved in each social action. The network approach also offers interesting possibilities for analysing the efficacy of policies. It is more useful to investigate the relationship structure that determines the employability of the migrants than to study employment numbers and characteristics. The latter depend on variables internal to the employment activity and on ever-changing external variables that generate a very uncertain information framework. The relationship structure, on the other hand, provides a more stable historical framework based on long-standing local relationships, regional cooperative capabilities, communication between both sides of industry in an economic district, and interaction between public and private agencies within a community.
China’s Evolution and Practice: Institutional Evaluation for Government - funded R&D

Professor Yan Fang
Director of Evaluation Department
China National Center for Science and Technology Evaluation, P.R. China
Email: fangyan@ncste.org

Delu Tian
China National Center for Science and Technology Evaluation, P.R. China

Evaluation of public sector R&D aimed at performance measurement is an integral part of the planning, monitoring and resource allocation. The role of Chinese government is shifting from conventional commander to service provider and regulator of macro social and economy. To buildup a service-centered, the accountability government has become a must. As main player in charge of allocation of R&D resource, the government has been paying attention to the problems of low efficiency in R&D programs to steer the supporting R&D in the overall social and economic developments. Thus, the institutional framework for evaluation has been development in China since 1994.

The paper attempts to describe and discuss the prevailing evolution and practices of institutional model for public sector R&D evaluation at the national levels with special reference to the organization of NCSTE - China National Center for Science and Technology Evaluation.

Firstly, the paper will briefly introduce the China National R&D programs and on which, how the government input, what the changing tendency in R&D expenditure annually and the sorts of output in R&D programs.

Secondly, the institutional model for R&D evaluation will be put forward and more issues will be discussed on the evaluation driving force, the governmental requirement, and evaluation culture as well as governance environment. In which, the organizing model such as evaluation procedure, professional training, evaluator developing and result reporting will be referred to participants with some of available cases.

The face of challenges in the institutional framework will be an attempt in presence in third part. The descriptions cite evaluation practice both in China and international experience to identify what are the critical and common challenges on institutional framework for public sector evaluation system.

The finally, the paper will raise some issues for discussion on future direction of R&D evaluation delivery, especially in developing countries.
Joining up public services - some implications for programme management and evaluation methods

Richard Breese
Programmes Manager
South Yorkshire Partnership, UK
Email: richardbreese@sypartnership.org.uk

I am a manager of regeneration programmes, interested in project and programme management methods. I am also in the final stages of drafting a Doctorate in Business Administration (DBA) thesis on joining up public services.

In my DBA thesis, I have devised a new framework for joining up public services which recognises the dimensional complexity of human activity, working from a critical realist standpoint. I will explain the implications of my research for evaluation, in terms of
- standardising evaluation methods across disciplinary boundaries
- the position of evaluation within the wider programme management process.

I will look critically at notions of causality, and how they are reflected in the approaches to both project/programme management and evaluation.
Establishing The System To Evaluate the Implementation of Social Protection Programmes in Slovenia

Dr Liljana Rihter
Assistant Lecturer and Researcher
University of Ljubljana, Slovenia
Email: liljana.rihter@fsd.uni-lj.si

Dr Nino Rode
Assistant Lecturer and Researcher
University of Ljubljana, Faculty of Social Work, Slovenia
Email: nino.rode@fsd.uni-lj.si

Barbara Kobal
Senior Researcher
Social Protection Institute of the Republic of Slovenia, Slovenia
Email: barbara.kobal@guest.arnes.si

According to the National Programme of Social Protection until 2005, the Slovenian state should enable the functioning of an integrated system of social protection. Information should be provided through an evaluation of social protection programmes. Therefore the project task of formulating an evaluation system of the implementation of social protection programmes was proposed by the Ministry of Labour, Family and Social Affairs.

The aim of the project, which will be described in this paper, was to establish an evaluation system which would enable a comparison of effects between similar programmes and to offer feedback to programme providers so as to allow them to better direct their future activity.

The first stage of the project involved identifying the starting points for the system evaluating social protection programmes. Yates’ evaluation model was used, according to which sources, procedures, processes and results of the programme need to be analysed and the relations between them evaluated. The existing, diverse social protection programmes were joined together to allow an evaluation using similar methods. Next, the evaluation criteria were prepared. The first complex includes general criteria based on the National Programme of Social Protection (quality of life, strength perspective, normalisation); the second complex includes special criteria which are defined in line with the findings of the profession regarding social problems; the third complex includes specific criteria formed within individual programmes for self-evaluation. Programme providers and users were invited to co-operate in the defining of special and specific evaluation criteria. On the basis of such criteria, measuring instruments were prepared to answer the question of whether the set aims were being achieved and what is the quality of work within a programme.

The last phase included pilot control of the implementation of evaluation in certain groups of programmes. The feasibility of the proposed evaluation method was thereby checked and the measuring instruments were tested.
Integrating ESF interventions in the Structural funds local development programmes

Dr Alessandra De Lellis
Researcher
ISFOL, Italy
Email: a.delellis@isfol.it

Co authors
Cristina Lion
Silvia Ciampi

The challenge of integration of interventions has been formulated in a stronger way than in the past in the 2000-2006 Structural funds programming. In this period, territorial integrated projects (PIT) and local development projects have been implemented with the aim to guarantee coherence between action financed by different financial instruments. Evaluation has faced these new forms of local development programmes, by elaborating methodologies and tools for the ex-ante, in itinere and ex-post evaluations.

This contribution presents the experience of evaluating integration within ESF programming, that is to say the role of human resources development projects in the context of local economies. This evaluation has been carried out by the National ESF Evaluation Unit in the framework of mid-term evaluation updating in the Italian Centre-north Regions.

Since the theme of integration has not fully been operationalised in the 2000-2006 ESF and ERDF programming, evaluation has firstly mapped the most relevant experiences in the Regional programmes and secondly ranked them according to different degrees of integration.

These different levels are linked to the potential synergy effects among projects, thus necessitating strong coordination activities.

The 2007-2013 programming stresses the integration principle as a way of implementing structural interventions. However mono-funding structures of programmes make it more difficult to integrate projects financed by different funds thus requiring increasing coordination. The main findings of the evaluation we discuss can support a better identification of integration ways.

Our discussion will focus on the challenge of evaluating integration, in particular the added value of this new principle which is considered by some policy makers difficult and expensive to implement.

Relevant and sustainable partnership, which is a building block to create social capital at the local level, is one of the components of this added value. Discussion is open to other components for the evaluation of integration.
The research sponsor, the program developer, his evaluator and her best practice list: conflicts of interest in the development, evaluation and dissemination of drug and violence prevention programs

Dr Dennis Gorman
Associate Professor
School of Rural Public Health
Texas A&M Health Science Center, USA
Email: gorman@srph.tamhsc.edu

Conflict of interest is becoming an increasingly important issue in the area of drug and violence prevention due to the way in which intervention programs are funded, evaluated, marketed and disseminated. Indeed, what is especially interesting about recent developments in the area of drug and violence prevention is that the field has sought to become more scientific in the conduct of program evaluations while at the same time becoming more business-like and commercial in the way that it markets and disseminates these programs. As with any product, consumer confidence is increased when an intervention is labeled “evidence-based” (especially by a seemingly independent party such as a government-sponsored “expert committee”), and thus to be “scientific” is to be marketable in the world of prevention practice.

This presentation will examine the implications of these trends in the evaluation and marketing of drug and violence prevention programs for the integrity of the research that is conducted, and specifically for the threats to research integrity that arise from conflicts of interest. It will begin by describing what is meant by the term conflict of interest and reviewing the main types of conflicts that have been discussed in the medical and public policy literatures in recent years. The remainder of the presentation will entail a detailed discussion of (1) the evidence indicating the extent of conflicts of interest in research into the efficacy of drug and violence prevention programs, and (2) the primary sources of individual, institutional, and political conflicts of interest in this area of research. It will be argued that in both the conduct of evaluation research and process of program dissemination there are sufficient potential individual, institutional and political conflicts that the research community should not only assess these but also put structures in place to ensure scientific integrity.
The evaluation of "integrated territorial projects": methodological challenges

Paulo Conceição
Assistant Professor
CITTA - Faculdade de Engenharia da Universidade do Porto, Portugal
Email: psc@fe.up.pt

Co authors
I Breda-Vázquez and P Moia

In the field of territorial or urban policies, both at national and European levels, references to "urban regeneration", "new urban policies", "integrated territorial projects" seem to configure new approaches to public policies. The corresponding initiatives, programmes and projects are based on concepts like:

- "integration", considering, in the same process, several dimensions (social, economical, physical, environmental) of urban problems and policies;
- "territorialisation", adopting specific ("area-based") targets and acknowledging the importance of spatial context for the definition of problems and opportunities of action;
- "participation", "partnership", "governance", constructing new ways of resources definition and mobilization and more complex institutional solutions (involving networks of agents, public and private, combining local, regional, national and supra-national levels of action);
- "learning", developing an experimental approach ("pilot-projects") and promoting the transferability of knowledge through the identification and dissemination of "good-practices".

The development of these initiatives, programmes and projects introduces two types of challenge to evaluation practices, the first one related to the role and institutional place of evaluation in these processes, and the second one related to the construction of methodologies and capacities to apprehend and evaluate the important dimensions of "integration", "territorialisation", "governance" and "learning". Both challenges require the understanding of relationships between evaluation practices and their - institutional and spatial - context.

The paper develops these questions and challenges, with a special attention to the Portuguese case. In the first part of the paper, a systematization of the main characteristics of recent urban programmes in Portugal is proposed. The second part of the paper identifies specific problems/needs, for evaluation practices, associated with these characteristics (the importance of references, in the evaluation process, to spatial vision; the attention to both procedural and "substantive" dimensions of the programmes; the attention to the ways, in specific contexts, concepts and dimensions of "integration", "territorialisation", "governance" and "learning" are constructed).

The third part of the paper discusses more general contributions, emerging from the study of the Portuguese case, to the debate on methodologies for the evaluation of "territorial integrated projects".
Evolutionary Evaluation

Professor William Trochim
Policy Analysis & Management
Cornell University, USA
Email : wmt1@cornell.edu

All evaluation can be understood as feedback within a dynamically evolving system. This paper outlines a general theory of evolutionary evaluation (EE), describes its epistemological and conceptual foundations, presents a simple systems model of EE and considers how evaluation methods can be understood within this framework. Philosophically, this work draws upon the evolutionary epistemology of Donald T. Campbell, and the foundational work of Sir Karl Popper, both of which suggest that ideas evolve in a manner directly analogous to all other evolving life and based upon the mechanisms of Blind Variation and Selective Retention (BVSR).

Evaluation is an integral part of the conceptual process of variation (BV) and plays a critical sensory role in the retention mechanism (SR). The EE model is based on a dynamic systems thinking framework where evaluation is viewed as an ongoing process that occurs simultaneously from multiple perspectives at multiple scales (nested part-whole hierarchies) over differing time cycles. The evaluand is embedded in an ecology of proximally similar entities, evolves through various natural stages (e.g., initiation, development, implementation, distribution; or birth, growth, maturity, death) and can be located within a phylogeny of systems phenomenology that ranges from simple to dynamic linear, dynamic and complex, and adaptive.

An EE approach has both practical and conceptual implications for evaluation. Conceptually it encompasses a broad range of traditional evaluation perspectives from ones that are more mechanistic, reductionistic, and realist to ones more dynamic, constructivist and holistic. EE offers a broad array of practical tools and heuristic devices that help transform the way we approach evaluation. Several tools and methods are presented including pathwork models, coupling methods, concept mapping, computer and social simulation, evaluation protocols, and informal rapid unobtrusive feedback processes. Issues of implementing EE and implications of the EE approach for evaluation are considered.
Expanding the Use of Theory to Improve the Future of Evaluation Practice

Dr Stewart Donaldson
Dean & Professor
Claremont Graduate University, USA
Email: stewart.donaldson@cgu.edu

Christina Christie
Claremont Graduate University, USA

This paper outlines strategies for increasing and improving the use of social science theory, evaluation theory, and program theory to evaluate programs and policies. Evidenced-based social science theories are often helpful for understanding the etiology of desired or undesired outcomes and for developing programs and policies for influencing those outcomes.

Evaluation theories are approaches or models that offer a set of prescriptions, guiding frameworks, specifying what a good evaluation is and how it should be done. One widespread variant of evaluation theory is centered on the notion of articulating and using program theory to shape those aspects of evaluation practice that are related to conceptualizing programs and policies, designing evaluation studies, and interpreting evaluation findings. Program theory and evaluation research based on it, in turn, are enriched when they can draw on relevant social science theory and when they contribute to knowledge about how social intervention brings about social change. This confluence of social science theory, evaluation theory, and program theory constitutes a distinctive approach to evaluation broadly cast as “program theory-driven evaluation science.” The potential of this integrative approach for improving the future of evaluation practice will be explored.
Clarifying evidence-based practice through program theory reconstruction

Tina Olsson
Doktorand
School of Public Administration
University of Gothenburg
Email: Tina.Olsson@gr.to

This paper uses the policy-scientific approach as described by Leeuw (2003) to reconstruct the program theories underlying three conceptualizations of evidence-based practice (EBP). The first two of these conceptualizations are taken directly from the literature and a third is taken from a relevant policy program: National support for knowledge development within the Swedish social services. This paper shows that “evidence-based practice” is not a homogeneous concept and interested parties within research, practice and policy may not have a shared vision of what EBP is even though they may use the same terminology. It is important for the successful implementation of EBP within practice and policy that a clear understanding of what EBP is and how EBP is supposed to work if research, policy and practice are to develop a shared understanding and work together toward common goals.
Creative tensions in commissioner and evaluator relationships

The Big Lottery Fund symposium

This will explore evaluation issues from the perspective of both the evaluator and the commissioner. The Big Lottery Fund has a substantial body of evaluation work, undertaken over the last ten years. Much of this was commissioned by BIG’s two predecessor organisations, the Community Fund and the New Opportunities Fund. Evaluations were typically conducted on a programme basis, and usually ran in parallel with the programme delivery. They were contracted out to external evaluators.

This symposium uses the experience of Big Lottery Fund Evaluation and Research staff, and that of contractors who have worked with us on major programme evaluations. The Healthy Living Centres, Green Spaces and Fairshare evaluations will provide particular examples, but other illustrations will be drawn from the 26 evaluations commissioned by CF, NOF and BIG.

We intend to use the symposium to explore some of the complex and occasionally conflicting areas that arise during the commissioning and implementation of long-term evaluations.

In particular we will explore:

- Tensions between generalist commissioners and specialist evaluators
- Tensions between specificity in evaluation design and the requirement for fully worked proposals, and the desirability of openness and responsiveness for evaluations to adapt to evolving circumstances
- Tensions in long-term contracts between commissioner and evaluator when the context, policy environment or organisational priorities change
- Tensions between inclusive participative design and evaluator professional boundaries; what is outside the commissioner’s or the evaluator’s brief?
- Tensions in expectations of commissioners to have early feedback (formative) and evaluators’ awareness of timescales for projects’ capacity to produce desired changes

The three speakers (Elliot Stern, Sally Downs and Sarah Mistry) will discuss these issues from the viewpoint of evaluators and commissioners and identify some particular challenges. There will then be a panel-led discussion on the issues raised.
Attending to the discursive nature of evaluation

Dr Jennifer Greene
University of Illinois, Urbana-Champaign, USA
Email : jcgreene@uiuc.edu

Professor Thomas Schwandt
University of Illinois, Urbana-Champaign, USA

Since the 1970s, largely as an outgrowth of the adoption of neo-liberal thinking in political-economic practices, the activity of evaluation has been increasingly associated with purposes of government and agency accountability, performance measurement, and outcome assessment. While these certainly are legitimate purposes for and uses of evaluation, they tend to portray the activity of evaluation largely as an instrumental activity involving the successful application of a set tools or procedures for determining outcomes or effects (a recent search on Google for the phrase ‘evaluation toolkit’ yielded over 15 million hits). Undeniably, evaluation activity involves the use of particular techniques, yet a preoccupation with technique too readily leads us to overlook the discursive nature of evaluation.

By "discursive" here we mean to emphasize two matters: First, that decisions about the value of programs and policies are typically constructed by stakeholders and evaluators in naturally occurring talk and interaction in a given practice (e.g., social welfare, education); second, that it is often via procedures for participation and dialogue within an evaluation that an evaluative conclusion is discursively reached through considering reasoned arguments and evidence.

Thus, this discursive character of evaluation draws our attention to important considerations in evaluation practice: One concerns the dispositions and virtues demanded of evaluators in the relational and interactional dimensions of evaluation work; the other concerns procedures whereby decisions about value are discursively reached in dialogue and discussion.

This paper will explicate and illustrate these dimensions of evaluation practice with the intent of helping evaluators to become more mindful and thoughtful about the discursive nature of evaluation and what that demands of them as professional practitioners.

This paper relates to the overall conference theme of examining the relational dimensions of evaluation as part of its role in contributing to democratic discussion about important social directions.
Action learning through evaluation partnerships: Private sector approaches to enhancing the mining sector’s contribution to poverty reduction

Kathryn McPhail
Principal
International Council on Mining and Metals
Email: kathryn.mcphail@icmm.com

Academic literature has focussed on a paradox whereby some countries with an abundance of natural resources have achieved less economic growth than countries without these natural resources -- the so called resource curse.

Yet mining and minerals have been instrumental in the development of successful economies and societies, where mineral wealth has not been associated with economic stagnation, conflict or corruption; and where prosperity and political stability now prevail.

Against this background, the International Council on Mining and Metals, which comprises 15 mining and metals companies, decided to launch an evaluation of the poverty reduction impact of mining investments at country level. The objective was to:

(i) identify the underlying success factors that explain why some countries have prospered from their mineral resource endowments while others have not; and

(ii) undertake analytical work in consultation with stakeholders in the sector: governments, labor organizations, Non Governmental Organizations, financial institutions, donor agencies and others - so as to seek ways forward and ensure better social and economic development outcomes from large-scale mining investments.

Work began in May 2004 and is now in it is third phase.

The paper will set out the process by which the evaluation work was carried out which involved considerable stakeholder involvement through two multi-stakeholder workshops, an international advisory group to ensure objectivity and rigor, a working group of 25 companies and chambers of mines from around the world, and two institutional partners: UNCTAD and the World Bank Group. It will also cover some key findings from the country evaluation case studies in terms of mining’s contribution to both economic growth and poverty reduction.

Finally it will set out some recommendations on how to enhance the contribution of mining to socio-economic development in future investments.
Exploring the possibilities of Critical Frame Analysis for evaluating policies

Dr María Bustelo
Associate Professor
Universidad Complutense de Madrid, Spain
Email: mbustelo@cps.ucm.es

Mieke Verloo
IWM- Vienna
Email: verloo@iwm.at

Some studies have pointed out that the type of policy evaluation less developed generally speaking is the so-called design evaluation, that based upon the content, focuses on the design or conceptualization of a policy, rather than on their implementation or results (Bustelo, 2003). However, this kind of evaluation is a key one, not only for the assessment of the coherence and relevance of policies, but also for offering a rationale for interpreting their implementation and results.

Studies on implementation problems of gender mainstreaming strategies suggest that they might be based on non-recognized and non-explored enough multiple interpretations and representations of the concept of gender (in)equality. Following this hypothesis, the MAGEEQ project (Mainstreaming Gender Equality in Europe) presents an approach at studying divergences in policy frames around gender equality as one of the elements connected to implementation problems (Verloo, 2004).

Based on the empirical and comparative research in the MAGEEQ research, this paper will explore the possibilities of Critical Frame Analysis as an important part of design or conceptualization policy evaluation, and will discuss how it can be crucial for the evaluation of gender policies and gender mainstreaming. The paper will further reflect upon the question if the use of Critical Frame Analysis can be extended to the evaluation and analysis of policies in general.

References:
Evaluation of the National Diabetes Support Team (NDST)

Miss Emily Kessler
Consultant
Matrix Research and Consultancy
Email: emily.kessler@matrixrcl.co.uk

The publication of the Diabetes National Service Framework (DNSF) in 2001 and the DNSF Delivery Strategy in 2003 established for the first time clear standards that diabetes services in England should be meeting, and a clear strategy for implementing them. The DNSF was a new way of delivering a national service framework (NSF). Whereas NSFs in the past had concentrated on strict targets and "top down" resource allocation, the DNSF was to be delivered across local organisations in networks through partnership working. The National Diabetes Support Team (NDST) was set up to help develop sustainable service improvement through aiding and supporting these partnerships.

Given that the NDST represents a new approach to supporting the implementation of an NSF it was imperative that it was supported by an action research evaluation to:

- deliver regular feedback to support the NDST's continued development and improvement; and
- consider the applicability of the NDST's approach to the implementing of future NSFs.

Matrix Research and Consultancy (Matrix) was commissioned by the NDST to evaluate their ethos and effectiveness over a period of two years. The overall approach to the evaluation is based on three high level research questions:

- Should it work?
- Can it work?
- Does it work?

The research has been conducted over a number of waves of fieldwork. The first examined the logic behind the creation of the NDST and identifying the 'theories of change' behind the intervention. Additional waves have taken a large-scale qualitative approach, interviewing those who the NDST have been working with to evaluate their operational success and impact.
Conceptualising the outcomes of community safety programs for natural hazards in Australia: Developing program theory using structured concept mapping

Dr Gerald Elsworth
Principal Researcher
Collaborative Institute for Research, Consulting and Learning in Evaluation
RMIT University, Australia
Email: gerald.elsworth@rmit.edu.au

In Australia, programs to increase community preparedness and self-reliance are an increasing feature of attempts to deal with the risk of natural hazards. Termed the ‘community safety’ approach, these programs represent a significant shift in emergency management thinking. In relation to wildfire, the programs support the implementation of the ‘Stay and defend – or go early’ household safety policy. Yet at present there is little understanding of how effective the programs are, for which communities and in what particular settings they work best, and how desired outcomes are generated.

The first part of the paper describes Trochim's structured concept mapping procedure and its use in eleven workshops across five Australian states to identify concepts that describe the changes or improvements community members and fire agency personnel believe are needed to make households and neighbourhoods safer from wildfire. Further statistical analysis enabled more detailed representations of the relationships between the identified concepts. A synthesis of individual workshop results yielded fifteen broad concepts.

A subsequent interpretative phase created two models of community safety outcomes that will provide the basis for subsequent program planning and evaluation. The first model, derived directly from the fifteen concepts, yielded an outcomes hierarchy for community safety programs. The second involved a reconceptualisation of the fifteen concepts and individual workshop results, yielding a representation of possible mechanisms and outcomes within a three-level (household, locality/community, and agency) view of the implementation context of community safety initiatives.

Recent theoretical work on free-hand concept mapping will be briefly reviewed. In Novak’s free-hand approach, concepts are typically represented in a hierarchical fashion with the most general and inclusive concepts at the top. Also concepts may be joined together by linking phrases to form ‘propositions’. The relevance for program theory development of these and other techniques from free-hand concept mapping will be explored.
Use of Evaluation Studies in the Eyes of Evaluators and Government Commissioners. Differences and Similarities between Polish and UK Patterns

Katarzyna Ekiert
Office of the Committee for European Integration, Polish Evaluation Society
Poland
Email: katarzyna.ekiert@pte.org.pl

Concept of this presentation comes from the paper Finding Out What Works. Understanding Complex, Community-based Initiatives, published by the King’s Fund, UK, in 2004. Its authors (A. Coote, J. Allen, D. Woodhead) examine there to what extent a number of flagship UK social programmes are really evidence based, what exactly role evaluation plays in those programmes’ management, and how evaluation studies help to build knowledge which over time improves policy and practice. Case studies, review of literature, seminar series and interviews with key persons commissioning, implementing and evaluating those governmental programmes were main data sources for the project.

As no similar research regarding evaluations commissioned by the Polish government has ever been carried out before, a paper presented by this abstract is an attempt to fill in that gap. Using the same research methods, it continues some universal issues raised in the King’s Fund study and provides a comparison between the patterns of evaluation use observed in both administrative and cultural environments. How is the role of evaluation is defined by the Polish government commissioners and evaluators? How empirical data derived from evaluation reports influence creating and carrying out policies? Why exactly evaluation studies are commissioned by the Polish central administration? What are the most popular evaluation methods? Why is it so? Are there any signs of a ‘learning culture’ built on evaluation research findings within central government in Poland? Bearing in mind undisputed differences between both a tradition and the current position of evaluation studies in the Polish and the British central administration, this study offers answers to the above mentioned questions.

Field research is constituted by 12 interviews with the Polish government commissioners and evaluators experienced in research on the large-scale governmental programmes carried out in Poland.

The author perceives her study as an introduction to more in-depth analysis concerning links between organizational culture, use of evaluation and a process of strengthening democracy as well as a contribution to testing some influential theories of evaluation use.
Filling the Gap between Evaluation Theory and Practice

Anneke Slob and Jan-Maarten de Vet
Coordinator Evaluation and Monitoring
ECORYS NL

An interactive workshop with theoreticians, practitioners, commissioners and users of evaluations

The initiators, Anneke Slob and Jan-Maarten de Vet, work as evaluation practitioners at the EU level and beyond. This proposal for an interactive workshop with different groups of actors is directly linked to the main conference theme “Critical Connections”. In their evaluation work they are confronted with a number of challenges that are insufficiently recognised, because the various stakeholders do not recognise the critical connections. Hence, the gap between evaluation theory and practice tends to become wider instead of smaller. As team leaders of complex evaluations the initiators are confronted with the following problems:

• A wide gap can be discerned between evaluation theory and evaluation practice. Evaluation theory has at its disposal a large and increasingly sophisticated range of methods and tools to be utilised under specific conditions; however evaluation practice requires often to ‘cut corners’ – as
  o Clients are not willing or able to put forward the resources, time and/or data resources to put theory in practice.
  o Moreover, theoretical sound evaluations are not always able to contribute to improved policy-making and are insufficiently geared towards the needs of policy makers.
  o The data constraints with which we operate generally do not allow strict application of the theory: the problem of lack of monitoring

• In practice, only basic methods (e.g. desk research, interviews, surveys) tend to be used in most of our evaluations and their utilisation is limited. Therefore, our evaluations are often less solid and crisp then possible – with often limited client excitement.

• Clients increasingly ask for clear and unambiguous conclusions and recommendations that can be readily implemented. However such conclusions can be easily contested, and therefore need to be evidence-based. This leads again to the need for stronger methodological foundations – that are not always sufficiently present in all of our evaluation work.

• New legislation makes evaluation and impact assessment extend to other areas than just spending programmes, including policy initiatives, legislation, agencies, etcetera. In these ‘untested waters’ there is much room for new methods and approaches and it is important to learn and exchange about such evaluations at the Group level.

• Although each commissioner has its own style, wishes and approaches, some clients such as the European Commission aim for a standard approach. From experiences elsewhere it is clear that such a standardisation approach is often followed by a diversification approach. After all a natural tension in evaluation work exists between ‘tailor-made’ approaches allowing to take into account specific wishes of the stakeholders and a standard approach allowing comparability between evaluations.

• A wide gap can be noticed between the wishes of the clients (as expressed by Terms of Reference and underlying expectations) and what can be delivered in practice. A strong methodology can play a big role in winning the work (with a weight of up to 40% of technical appraisal) and subsequently managing these expectations.

The initiators have started a research project in their own working environment, the ECORYS Group which has worldwide experience in a large number of evaluations, to address the challenge of filling the gap. One of the aims of the research project is to develop a practical toolbox and to use this as practical guidance in the daily evaluation work. ECORYS is working in close collaboration with evaluation theoreticians to improve the quality of the work. However, ‘filling the gap’ remains a major challenge.
Supervision and work-related well-being - possible effects and outcomes of supervision in the welfare services.

Professor Synnöve Karvinen-Niinikoski
Department of Social Policy Studies,
University of Helsinki, Finland
Email: Synnove.Karvinen@helsinki.fi

Supervision in the case of social professions is used to improve professional practice and support the work-related well-being. This paper ponders how to evaluate the role and possible outcomes and effects of supervision. Welfare and health services and especially social work organisations are considered to be in crises when the supply and devotion of qualified staff is regarded. The concern is about ill-being and non-motivating career prospects and lacking possibilities for developing professional expertise that are causing problems to recruit and motivate devoted and qualified professionals to stay in the field.

Supervision is one of the central methods in supporting the work-related well-being and in relieving ill-being. It is used in various fields with strong emphases on offering the social and psychical support. This supervisory field is very busy and a professional expansion towards supervision-business is going on. However, very little about the possible outcomes and/or effects of supervision are known. Still supervision can be seen as a remarkable investment with both economic as well social and health political effects. A joint feature in the supervisory approaches to work-related well-being is the need for employees to become conscious of their own feelings and experiences as factors affecting their work-related well-being and to learn to understand and to see the interconnectedness of these experiences in the organisational dynamics and the development and changes in the working life. All this is difficult to be measured and evaluated.

Does supervision as practiced today within the Scandinavian countries meet any of these needs and how? With what kind of an approach and methods supervisory practices could be evaluated? This paper will discuss the possible options for conducting this kind of evaluation.
Evaluation as a Tool of Change Management: critical reflections from a cross-discipline/cross-profession evaluation

Dr Peter Wells
Principal Research Fellow
Centre for Regional Economic and Social Research
Sheffield Hallam University
Email: p.wells@shu.ac.uk

Dr Bea Jefferson
Yorkshire Forward, UK

This paper reports on the interim findings of a study commissioned by Yorkshire Forward, the Regional Development Agency for the Yorkshire and Humber region in the United Kingdom. In 2005 Yorkshire Forward commissioned a consortium of evaluation professionals, academics and economic development specialists to undertake a project entitled Innovative Evaluation Methods. The project consists of three project evaluations which will trial innovative evaluation methods and a meta-evaluation of the three project evaluations. The projects which are being evaluated and the evaluation itself are seen as a process of change management through which it is intended that evaluation:

- Will become seen as part of a learning and improvement process
- Create forums for reflective practice through action learning
- Develop a toolkit to guide project managers and evaluators in the future
- Demonstrate the role of evaluation in informing organisational strategy and policy

The individual evaluation projects are also working with three distinct groups of practitioners and professionals: with Education Primary and Secondary School Teachers (in the case of a schools enterprise project); with economic development agencies and organisations (in the case of a cluster champion and a 'customer first' business support project); and with research and evaluation professionals (in the case of the meta evaluation). One of the key purposes of the study with each group is to change and improve practice.

The hypothesis of the evaluation project is that the use of innovative evaluation methodologies can improve the utilisation and capacity of evaluation to improve practice. This is undertaken by a meta-evaluation which both acts to synthesise the evaluations but also to provide a critical challenge through drawing on a series of evidence reviews and from an action learning set of research and policy practitioners.

The paper will explore specific issues around the relationship between evaluation and the processes of change management in different professional, political and institutional contexts. It will examine the reception of evaluation in these different contexts the implications this holds for future cross-discipline/cross-profession evaluations.
Evaluating career tailoring and succession planning in General Practice for policy and practice

Dr Loretta Bellman
Independent Consultant Nursing and Healthcare
Lambeth Primary CareTrust/Guy's, King's & St Thomas' School of Medicine, UK
Email: loretta.bellman@btinternet.com

In a time of considerable change and uncertainty in the NHS it is unclear how the provision of primary care will develop in the future. Retaining and supporting the older NHS workforce, particularly in general practice, is an important, yet controversial, factor. Support and development for general practitioners is generally aimed at the younger doctor. However, reports of a declining interest in general practice, and changes in the career paths of young general practitioners, present a possible challenge regarding future numbers of doctors available to meet the needs of a primary care led NHS.

There is also a need to retain ‘wise heads’ and value their contribution. Yet, current literature points to a future retention crisis amongst older general practitioners. This is due to a number of factors, including: the aging population of GPs in the UK; the current unremitting pressure of work; relentless change; lack of job satisfaction; concern regarding new technology, and views of disempowering national policies.

There is now national and local recognition of the need to retain and support the aging workforce in general practice. Contemporary literature, together with stakeholder involvement, have provided both the evidence base for, and the development, implementation and evaluation of an innovative career tailoring and succession planning project in SE London. A key outcome of the project is to develop a Primary Care Trust (PCT) strategy for career development for older doctors that can be shared with other PCTs both locally and nationally.

A participatory evaluation is exploring stakeholders’ experiences of the evolving project, identifying actual and perceived outcomes from participating in the project, and aims to highlight models of good practice. The presentation will provide an overview of the project and explore the extent to which the evaluation has informed local policy and advanced peri-retirement opportunities.
Evaluating Social and Environmental Aspects of Sustainability Together – Ways to Transcend the Discipline Divide

Dr Colin Hocking
Senior Lecturer in Ecology and Sustainability
Victoria University St Albans campus, Australia
Email: colin.hocking@vu.edu.au

Several large scale projects funded jointly by State Government, private organisations and not for profit foundations, are under way in Victoria, Australia, to jointly foster environmental and social outcomes in urban communities. There is urgent community based need for action to reverse environmentally damaging behaviours, and at the same time maintain community resilience that will allow adaptation to peak oil (and increasing fuel prices), climate change effects, severe water restrictions and a wide range of other impacts. Opportunities exist in the projects to investigate the inter-relationships between social strengthening (community building) and behavioural change for sustainability (reduced waste, water use, greenhouse gas emissions).

Methods to achieve these outcomes are drawn from a range of disciplines, including environmental project management, social psychology, community development, advocacy, personal development, and communications practice. The documentation and evaluation of the social and environmental outcomes of the projects, and their interactions require that a combination of methods are used, from the fully quantitative to the fully descriptive, and a range of techniques in between.

A useful way to frame evaluation in the project has been to identify underlying agreed pedagogies for action and evaluation, across participants, funders and other stakeholders. This has centred around flexible forms of action research, action learning and dynamic capacity building, with the evaluator employing constructivist approaches, and acting as much as critical friend and process facilitator, as documenter and interpreter. Ways in which stakeholders with primary reference frames arising from different disciplines (scientific, social, psychological, political, communications, etc.) can be facilitated to orient to a range of common purposes with objectives and evaluable outcomes that transcend disciplines are discussed in the analysis of this work.

Examples of how to relate the various types of outcome to one another, and interwoven to provide a scaffolding of cross-referenced interpretation are also discussed. Transdisciplinary evaluation will increasingly be required for large scale community sustainability project in future, and development of appropriate techniques is urgently needed.
Evaluation as Facilitator in Safety Prevention

Chris Chilvers
Evaluation
Suva, Swiss National Accident Insurance Fund, Switzerland
Email: chris.chilvers@suva.ch

Evaluative knowledge helps Program Managers and Policy Makers improve their understanding in their field of activities and decision making. It also helps users express their opinion and develop consensus among the actors.

A Safety Program in the high-risk building branch in the Italian part of Switzerland serves as illustration.

After tough salary negotiations between workers and employers (with demonstrations and boycotts) a common safety campaign with the same actors had thereafter little chance to succeed.

This contribution explores how an evaluation of the safety standards and safety knowledge on the building sites helped to turn this conflict situation into a successful win-win campaign. It also illustrates how evaluation can be an ideal guiding instrument and learning process for all involved and at the same time contribute to building up consensus and understanding among the actors. The role of evaluation as an independent service or as part of the prevention campaign itself will also be discussed.
Critical connection between evaluation and decision-making: the example of an evaluation mechanism of EU policies in the area of Freedom, Security and Justice

Dora Correia
European Commission
DG Justice, Liberty and Security

Evaluation in the European Commission has developed since the 1980s. Although about 170 evaluations are carried out yearly, there are some challenges ahead. Amongst them is the critical connection between evaluation and decision-making.

This is partly due to the missing link between evaluation of policy instruments (be they financial or regulatory) and the high level political decision level. Thus, a mechanism to bring together evaluation results of individual instruments; assess their relative contribution to the overall policy objectives; and the level of achievement of the policy as such is necessary.

Such a mechanism should include:
- high quality data gathering,
- overarching and political synthesis of evaluation results,
- fostering an evaluation culture at the different layers of governance
- wider dissemination of evaluation conclusions and recommendations.

In this context, the European Commission has proposed to set up a coherent and comprehensive evaluation mechanism of EU policies in the Justice, Freedom and Security area.

The proposed mechanism is threefold and progressive:
- First, it provides for setting up a system of information gathering and sharing.
- Second, it includes a reporting mechanism which consolidates, uses and analyses this information, via the presentation every two years of a general evaluation report.
- Third, it is completed by targeted in-depth strategic evaluations.

Such a mechanism would build on current practices and use information resulting from the application of existing evaluation requirements. It will be comprehensive and encompass all the policies in the area of Freedom, Security and Justice.

This mechanism is expected to (1) further improve the knowledge base of policy-making, by identifying problems and obstacles encountered when implementing policies, programmes and regulatory instruments, as well as their relative effectiveness, (2) contribute to the financial accountability and transparency of policies, (3) favour learning and an exchange of good practices and (4) participate to the development of a culture of evaluation across the Union.

The present mechanism is under development and could be applied to other areas. The presentation will focus on how it will contribute to improving the critical connection between evaluation and decision-making, and on transferable practices to other contexts and areas.
Evaluation Roots in the USA and in Europe: tracing traditions?

Dr Ove Karlsson Vestman
Professor
Malardalen university, ISB, Sweden
Email: ove.karlsson@mdh.se

Dr Marvin Alkin
Professor
UCLA, USA
Email: alkin@gseis.ucla.edu

Dr Frans Leeuw
Professor
Wetenschappelijk Onderzoek & Documentatiecentrum. WODC,
Ministerie van Justitie, The Netherlands
Email: f.leeuw@minjus.nl

Dr Nicoletta Stame
Professor
University of Roma "La Aapienza", Italy
Email: nicoletta.stame@uniroma1.it

Dr Reinhard Stockmann
Professor
Universität des Saarlandes, Germany
Email: r.stockmann@mx.uni-saarland.de

This panel will present perspectives on the Evaluation Theory Tree published in Evaluation Roots as a stimulus for discussion of evaluation theoretical perspectives of European evaluators. The Evaluation Theory Tree posits three aspects of evaluation, which are the most dominant in particular evaluation approaches: use, methods, valuing. Obviously there may be substantial variation in the various presentations by European scholars who, probably, will present several other possible perspectives on evaluation as it takes place in a European context.
Evaluation in Developing Countries panel session

DFID’s Evaluation strategies in the context of a strengthened commitment to the Elimination of World Poverty

Chair:
Colin Jacobs, British Council

Presenters:
Stephen Lister, lead author of the recent joint General Budget Support evaluation
Julian Barr, consultant (ITAD), Country Programme Evaluations

The new DFID White Paper (Eliminating World Poverty: Making Governance Work for the Poor, 2006) sets out an agenda for expanding the UK aid programme in the period up to 2010 and beyond. This period will see development expenditure rise, and the level and expectation of evidence of ‘delivery’ and impact attributable to DFID expenditure rising correspondingly. External scrutiny about performance and achievement of results will continue to intensify. Additionally, in its recent (2006) Peer Review of the UK, the OECD Development Assistance Committee (DAC) welcomed considerable progress in the strengthening of DFID’s evaluation capacities and outputs, but has also called for reinforced monitoring, evaluation and knowledge management systems.

Thus, along with the rest of the donor community, DFID is introducing a results-based monitoring and evaluation framework in all aspects of its work. Higher priority is being given to the systematic assessment of the outcomes achieved from the policies and programmes. DFID also faces the same evaluation challenges confronting both donors and aid recipients in the twenty-first century:

- aid interventions are increasingly large-scale, and undertaken on a joint basis, and yet attribution is considered increasingly important;
- both bilateral and joint design and implementation processes are long and complex, and yet the desire for an early understanding of impact and the success (or otherwise) of development outcomes is strengthening;
- both donors and partner countries emphasise partnership and the importance of country ownership, and yet much of the historical evaluation capacity is held in the developed world.

DFID’s evaluation department is working across the organisation in a variety of different ways to strengthen understanding of impact, and to strengthen our internal capacity to evaluate. We are also committed increasingly to working jointly – as we work to implement the commitments put forward in the 2005 Paris Declaration on Aid Effectiveness – with both partner countries, and with the agencies and organisations participating in DAC networks.

The session will outline some of the key ways in which DFID is responding to these challenges. Three main areas will be presented:

1.) A recently-prepared paper outlining DFID strategy for improved impact evaluation. Over recent years, the UK Government has considerably increased the volume of aid given to developing countries, and has sought to move from aid administration, to the promotion of coherent development policy. This has meant not only understanding the outcomes of bilateral development interventions, but also working internationally to learn and apply lessons across different continents in different institutional contexts. A significant increase in the volume of aid has at the same time increased the need to account for its effectiveness, both to partner countries, and to the UK constituency (including especially politicians), and to develop a robust understanding of DFID’s success (or otherwise) in
meeting its ambitious poverty reduction goals. Simultaneously, there has been increasing interest in the ‘evaluation gap’ and a perceived need for the rigour of increased impact evaluation. DFID is currently developing an internal policy on impact evaluation. This will be presented in brief, within the context of the recent Bellagio meeting on Impact Evaluation, organised by the Centre for Global Development, and forthcoming efforts to reach a common position with the development assistance committee of the OECD.

2) A summary of the findings and challenges of the recent joint evaluation of General Budget Support, which DFID led on behalf of 24 donor agencies, and 7 partner countries (presented by Stephen Lister, principal author of the Report). The joint study undertook to evaluate the development effects and lessons to be learned from the provision of ‘budget support’ by 24 agencies over ten years, in 7 different country contexts. A logic model was developed and used for this evaluation, potentially taking the method to its limits. A logical framework was designed to enable the intervention logic to be spelt out and the most important component parts tested. The evaluation faced the further challenge of avoiding the imposition of a mechanistic paradigm on an organic situation. The evaluation was led by Stephen Lister, who will present the findings in brief, and discuss the challenges of extending the original evaluation design, and the experience of undertaking and synthesising the work.

3.) An outline of the findings and methods of a rolling programme of Country Programme Evaluations, undertaken as ‘light-touch’ evaluations to provide both internal accountability, and useful information and perspectives for the country offices themselves (presented by Julian Barr, ITAD, the consultancy group which manages this intensive programme of evaluations). The first synthesis of findings has recently been concluded, and the methodology for the studies refined.
Evaluating training effects, its connection with daily practice (the International Training Centre of the ILO as illustrative case)

Dr Frans Lenglet
Director of Training
International Training Centre of the ILO, Italia
Email: f.lenglet@itcilo.org

The presentation will focus on how in the daily operations of an international training institution with 450 activities and 11,000 participants annually, the various types of training and learning "effect" receive the attention they deserve, in a cost-effective way, while satisfying many different stakeholders.

The first section gives the background against which the evaluative practice has been developing over the years. It will discuss the internal need for feedback on the relevance and effectiveness of training products services as well as the need for continuous improvement. Will also be discussed the requirements of donors and sponsors to conduct a summative evaluation of training projects and interventions, as well as the political need to visualize and justify the training institution's financial and other efforts.

The second section describes the practice and the instruments of evaluating training programmes and events, concentrating on outputs (numbers and characteristics of participants, participants' reactions and satisfaction, immediate learning effects), outcomes (individual performance) and impact (organisational and institutional performance). It will highlight "new connections" by discussing the role of ex-ante evaluation and cohort evaluation.

The third section discusses seven (problematic) aspects of evaluating training effects: attribution, impact wave, distance, institutional resistance, power of influence, group cohesion and traceability. When dealt with properly, attention to these aspects could enhance the value and quality of the evaluative practice. It also could make it more complex and more costly.

The fourth section places the issues discussed earlier in the wider context of result-based management. It discusses the extent to which full-fledged evaluations, encompassing all types and levels of effect, are useful or necessary for managing a training institution.

The final section makes suggestions about how the content of the discussion on evaluative practice could be transformed into a comprehensive training curriculum for trainers and training managers and other decision makers concerned with making training as relevant and effective as possible within real world constraints of time and resources.
Impact Assessment of Training for companies development in the Russian Federation

Luca Quaratino
ISTUD – Senior Researcher
IULM – Researcher in Organizational Science, Italy
Email: luca_quaratino@istud.it

Caterina Carroli
ISTUD – International Projects Department - Junior Researcher, Italy
Email: carroli@istud.it

Luigi Serio
ISTUD – Director of Institutional Projects Department, Italy
Email: lserio@istud.it

The paper addresses the issue of designing and realising an impact assessment study of management training in a transition economy, i.e. the Russian Federation.

It is based on a project carried out between January 2005 and April 2006 in the Russian Federation, notably in the areas of St. Petersburg, Nizhny Novgorod and Tver. The purpose of the assessment study was to analyse the long term impacts of national and international training projects on management development, and to share the results of the study with the principal national and international stakeholders in order to support their future policymaking (European Union, World Bank, Russian Government. A total of 179 interviews were effected in the three regions – addressed both to company managers and to a wider target of significant stakeholders (training providers, public administration, industrial bodies and representatives of Alumni Associations). The total number of companies studied during the impact assessment were 30. The typologies of the projects included in the sample were the most significant training programmes addressed to Russian private companies funded by national and international donors.

From an epistemological point of view, the decision was to adopt an interpretative approach (Guba, Lincoln, 1981), using the methodology of ‘multiple case studies’, qualitative methods and “field work” as essential parts of the study (individual in-depth interviews, group interviews, focus groups on specific targets, participant observation and document analysis).

The paper discusses:

• the methodological and “political” issues faced during the study coming from the adoption of such a methodology (managing the relation with the different stakeholder of the projects, overcoming language and culture diversities and their impact on the evaluation process);
• the main findings of the study, describing the impacts of management training at individual, organisational and ‘local environment’ level;
• the most significant recommendations emerging from the research that were presented and discussed with the national and international policymakers;
• some lessons learned, which might represent a useful “shared knowledge” on designing and implementing management training projects in developing countries and transition economies.

1 The European Union invited the European Training Foundation (ETF) to undertake this task. In order to carry out the implementation phase, the ETF enlisted the support of an independent organisation with expertise in the field of management training evaluation, namely ISTUD – Istituto Studi Direzionali.
Monitoring and Evaluation of Research and Development Projects in Developing Countries Committed to Poverty Alleviation

Dr Carlos Moreno
Senior Research Fellow
Natural Resources Institute, University of Greenwich, UK
Email: C.J.Moreno@gre.ac.uk

We propose a symposium session which will explore the challenges of Monitoring and Evaluation of research and development projects in the context of developing countries committed to poverty alleviation. With the rapidly changing context of globalisation and debates on the relationship between economic development, poverty and democracy, evaluation becomes highly significant in its potential to influence in-country public policy and government, donor and international efforts to achieve the millennium development goals. The focus of this session is on how monitoring and evaluation can contribute to improving the integration of policy formulation and its effective implementation, by involving a range of stakeholders, by incorporating diverse perspectives, by exploring the institutional dimensions which influence policy dimensions, alongside the use of emerging approaches (e.g. innovation systems thinking, knowledge networks, learning alliances, multi-stakeholders process and outcome mapping).

Presentations:

EVALUATION IN DEVELOPING COUNTRIES IN SUB-SAHARA AFRICA (SSA)

Mike Morris
Natural Resources Institute, University of Greenwich, UK

Social development is a complex process dependent on a system-wide constellation of players and the interactions between them. If interventions are to be sustained and lead to transformational outcomes, then complex changes need to take place across the whole system, and become embedded in the respective (i.e. local, through national) institutional settings.

However, interventions often operate in an institutional vacuum, failing or being unable to take adequate account of the institutional setting (e.g. market failure, weak state institutions, unfavourable policies, lack of capacity, conflicts of interest). Capacities created within project cycles are either not scaled-up or not suitable for scaling-up, and are lost after the project’s conclusion. Up-take - and therefore impact - is thwarted by unforeseen factors. These factors may be related to household circumstances (e.g. vulnerability, lack of voice, limited resource portfolio, and poor access to markets) or to wider institutional issues (e.g. poor input markets, high transaction costs, inadequate staff incentives, information barriers, ICT skews between levels, product registration procedures).

Donors continue to place considerable emphasis on developing country governments establishing pro-poor policies and legal frameworks and many policy documents can now be discovered on state websites eloquently linking sectoral policy with PRSPs and 2020 visions. Policy formulation processes tend to remain within the domain of senior ministry staff, with consequent limitations for implementation strategies. Evaluation of the impact of policy is done through assessment of the sector’s performance as a whole, with policy implementation seldom monitored. Livelihood patterns, the mix and access to resources (land, water, services etc) of the majority of people in many SSA countries, remain however dependent on informal institutional arrangements (e.g. customary or religious laws). Formalisation of the institutional arrangements is strongly linked to the overall
development of national economies, such that most sectors in SSA, developing countries remain predominantly informal, and policies, laws and administrations have limited reach.

Addressing these broader institutional issues and impacting on poverty requires a different approach from that of conventional relationships between the state, voluntary and private sectors, which are often characterised by distrust, fragmentation and short-termism. Various emerging approaches (e.g. innovation system thinking, knowledge networks, learning alliances, multi-stakeholder processes, outcome mapping etc) would appear to provide space for stakeholders to explore alternative strategies for dealing with the risk and uncertainty in these predominantly informal systems, and for the adoption of more socio-intensive approaches.

This paper attempts to capture the wide range of relationships that will have to be nurtured to fulfil the potential of evaluation in democracy and development as well as some of methodological approaches for doing it.

**ASSESSING THE SOCIAL IMPACT OF CODES OF PRACTICE: EXAMPLES FROM THE CUT FLOWER INDUSTRY, KENYA AND THE WINE INDUSTRY, SOUTH AFRICA**

Adrienne Martin and Valerie Nelson.

In the last decade, increasing attention has been paid globally to the ethical sourcing of commodities and to the social and environmental conditions in export industries in developing countries. Various internationally recognised and local codes of conduct and practice have been developed, promoting environmentally, and socially responsible standards of business. This study was an attempt to explore systematically the impacts of codes of practice on those whom they were intended to benefit – workers, workers households and other local stakeholder groups, with a view to informing policy makers, donors and industry bodies.

Two examples of the application of codes of practice were studied in detail; in the cut flower industry, Kenya and the Western Cape wine industry, South Africa. The study aimed to provide a longitudinal assessment of the differences made to welfare and livelihoods by the introduction of codes of practice, based on comparisons of conditions for workers across code-adopting and non-adopting companies, using a combination of qualitative and quantitative methods. The main dimensions of impact examined were - material wealth, social well-being and human capital measures and worker empowerment.

The paper outlines some of the features of the impact assessment methodology, the findings and the challenges presented; in particular the problems of information access, securing participation, continuity in a longitudinal study, and issues of attribution in a rapidly changing complex context both nationally and in terms of international trade relations.

**LESSONS LEARNT FROM THE SOCIO-ECONOMIC IMPACT MONITORING OF A SCIENTIFIC PROJECT ORIENTED TOWARDS POVERTY REDUCTION THROUGH THE DEVELOPMENT OF THREE OPEN-POLLINATED SEED VARIETIES RESISTANT TO TOMATO LEAF CURL VIRUS (TOLCV) IN SOUTH INDIA**

Carlos Moreno

Co authors


Since 1996, the University of Agricultural Sciences in Bangalore (UASB), India, the Natural Resources Institute (NRI) of the University of Greenwich, United Kingdom (UK) and the Asian Vegetable Research and Development Centre, (AVRDC), Taiwan, have been conducting scientific research in
India, to understand the sustainable management of ToLCV disease and its whitefly vector, Bemisia tabaci (Genn.). This virus has a dramatic effect on farmers’ ability to grow tomatoes and has seriously affected their livelihoods for almost two decades. In fact, the whitefly and the plant viruses it transmits, have achieved worldwide prominence, due to the enormous yield losses they cause in tropical and sub-tropical crops such as cotton, cassava and many vegetables, including tomato. The project’s objective was to find a suitable solution to assist farmers to grow tomatoes in the south of India.

In 2002, after 6 years of scientific research, three new types of virus-resistant open-pollinated seeds were developed by the partnership of the UASB, NRI and AVRC. Later, in the same year, and after official approval of the Indian Government, the three resistant seeds – Nandi, Sankranthi and Vybhav were released commercially. This was one of the first university scientific projects within the UASB to venture into the marketing of a technology to various stakeholders, including the private sector. The paper presents the interesting trajectory of the project which was initiated as scientific research funded through a poverty-oriented donor research programme and culminated in successful marketing of the seeds. As part of that process, various stakeholders such as farmers, researchers and private companies participated and took a keen interest at various stages of the project. The discussion covers the monitoring approach used by the project to assess impact and, secondly, some of main results and lessons learnt along the way.
Evaluating Measures against Right-Wing Extremism in Switzerland

Christian Hirschi
University of Zurich
Department of Political Science, Switzerland
Email: chirschi@pwi.unizh.ch

Co author
Thomas Widmer

In this paper we demonstrate the specific characteristics that have to be taken into account when evaluating measures against right-wing extremism. Focusing on right-wing extremism and corresponding countermeasures in Switzerland, we address the following questions: 1) What are the measures taken against right-wing extremism? 2) What measures work best against right-wing extremism in Switzerland?

As in many European countries, right-wing extremism became an issue again in Switzerland over the last decade. Most of the time a latent phenomenon, right-wing extremism appears on the political agenda as soon as a constituency is confronted with manifest rightist incidents such as reunions of right-wing extremists or racist assaults of individuals or groups. But what kind of countermeasures to be taken often remains unclear since empirically based evidence about the effectiveness of specific measures is widely lacking. Evaluation may contribute to a more systematic understanding of the effectiveness of specific measures against right-wing extremism. In our paper we show that well-defined measures with specific goals and quantifiable anticipated effects on specific addressees are usually more effective than loosely formulated measures targeted at the general public to control the spread of right-wing extremism in all its facets.

The paper is based on the findings of a larger research project on right-wing extremism in Switzerland and draws on an evaluation synthesis of twelve evaluations of political measures taken against right-wing extremism in Switzerland, Germany, the United States, England, Australia, and the European Union and additional evaluations of eight measures on the Swiss federal, cantonal, and local level in the fields of social work, youth work, education, law and order, public administration, and media.
Evaluating Programs for Children

Dr Shannan McNair
Associate Professor
Oakland University, USA
Email: mcnair@oakland.edu

Evaluation across “contexts” includes the contexts of age and program type. Youth programs included public and private institutions, out-of-school learning and youth development experiences, interventions for health awareness and prevention of substance abuse, and parent/child interventions. Programs of all kinds for children engage evaluators in challenges that are sometimes similar, but often different from those confronted in evaluating adult programs. This session will engage participants in an overview of challenges and opportunities common to evaluating children’s programs. The development of conceptual models will be explained, and appropriate instruments and strategies for various age groups and program types will be discussed.

Participants will be engaged in brief activities/application of ideas, problem-solving and discussion. Practical examples will be provided.
The Validity, Utility, Credibility, Cost-Effectiveness, and Ethicality of Research Evaluation Systems: A Comparative Analysis of Seventeen National-Level Systems

Chris L. S. Coryn
Research Associate
The Evaluation Center - Western Michigan University, USA
Email: christian.coryn@wmich.edu

Worldwide, under demands for greater accountability and diminished funding, many countries have initiated systems for evaluating publicly-funded research at the national level. The evaluation of publicly-funded research now has a long tradition, particularly in the European countries, dating to the early 1970s.

Although there are vast differences in the way governments fund research around the world, and a diversity of approaches to evaluating publicly-funded research, almost all now share the common purpose of relating funding to performance. Large-scale research evaluation systems have now become commonplace throughout the world to evaluate a nation’s researchers, their research, and performance between and within research fields and disciplines, institutions, research groups, and in comparison to international norms and standards, for instance.

The purposes of these national-level systems include, among others, priority setting, allocation of research funding, improvement, competition, and decision making in general. The nations and research evaluation systems included in this analysis are Australia’s RQF, Belgium, the Czech Republic, Finland, France, Germany, Hong Kong’s RAE, Hungary, Ireland, Japan, the Netherlands’ SEP, New Zealand’s PBRF, Poland, Sweden, Taiwan, the United Kingdom’s RAE, and the United States. This paper begins by presenting broad accounts of the context in which national-level research evaluation takes place, and their purposes and methodologies as developed and employed in more than a dozen nations and regions throughout the world.

The paper concludes with a comparative analysis of these systems applying criteria of validity, utility, credibility, cost-effectiveness, and ethicality. The general analytic approach used is a pragmatic one, and is not intended as an attempt at a complete classification or generalizations, as the systems themselves are heterogeneous, and any such attempt would easily fall into the trap of simplification and therefore quickly become false.
A Selection of Criteria to Metaevaluate University Evaluation Systems

Dr Baltasar Fernández-Ramírez
Lecturer on Social Psychology and Program Evaluation
University of Almería, Spain
Email: bfernandez@ual.es

Co authors
Enrique Reboloso
Pilar Cantón

The rhetoric of quality has permanently entered the European university system. Evaluation has become an indispensable instrument for managing university academic and administrative affairs. European convergence positions evaluation as one of the pillars of convergence: the quality guarantee of institutions, degrees and educational systems, as an argument for facilitating the processes of recognition and student and professional personnel mobility.

The four-stages evaluation model has become established as a paradigm of good practice. The variations are, however, important: What are the goals pursued by evaluation? How central is it to decision-making? How is its influence promoted in the improvement and internationalisation of the system? In line with some European initiatives, we need to define and test procedures for metaevaluation and quality assurance of the evaluation processes themselves. The criteria available are many (international professional standards, ethical principles, scientific standards, criteria related to specific evaluation activities).

The basic problem is deciding what metacriteria are to be used to carry out the selection of the criteria applicable to our case. That is, to answer the question of what kind of evaluation we need to make directed at what kind of university. In this article, we describe the decisions made to define a first set of criteria applicable in the metaevaluation of university quality evaluation processes.

The structure of criteria, phases and matters prior to the evaluation which is the subject of analysis is briefly described. The metacriteria used are also discussed and their value in the evaluation processes is analysed in a first sample of participant opinions.
Evaluation of education and research at Universities: critical connections in the Italian experience

Marco Costantini
University of Chieti, Italy

Co author
Dr Emanuela Reale
Researcher
CERS CNR, Rome, Italy
Email: e.reale@ceris.cnr.it

During the nineties the use of evaluation as a tool for driving the allocation of public resources to research performers was introduced in almost all the European countries both at ex ante (for project funding allocation) and ex post level (for block grant allocation).

One of the most important critical connection for the ex post evaluation is how to harmonise different methods and instruments for judging the quality of kind of institutions, such as Universities, where different functions co-exist, more specifically education and research.

In fact, the evaluation of educational activities generally use standards related to the expected performance, which represent the minimum requirements for guaranteeing the quality of the service provided. The standards are operationalised through a set of quantitative measures on the outcome of the educational processes, the efficient use of the available resource and the effectiveness of the educational programmes.

On the contrary, for evaluating research it is important to take into account assessment methods and instruments aimed to judge the scientific quality of the research products in term of capability to produce significant breakthroughs of the scientific knowledge as well as to generate an impact in socio-economic terms.

The paper aim is to analyse some critical connections for evaluating education and research within Universities. Main questions to be answered are: what does quality means for the different evaluation purposes, how can we harmonise the results deriving from the two evaluations processes, what is the way for governing the evaluation of education and research within the academic institutions, how the different evaluation results impact the Government resource allocation.

The analysis developed by making reference to the results of the Italian evaluation experiences, carried out by the two national Committees in charge for the assessment of education (CNVSU) and of research (CIVR) in the last five years (1999-2003).
Evidence of Worth in Not-for-Profit Organisations

Ian Harris
Director
Z/Yen Limited
Email: ian_harris@zyen.com

Not-for-profit organisations are coming under increasing scrutiny. Their profile is growing as they deliver more public services and employ more people (employment in the voluntary sector is growing by 10,000 each year and currently stands at 608,000. Income from service delivery makes up almost half of not-for-profit sector income or £12.4 billion per annum). As a result they face growing demands to evaluate their activities and provide evidence that their work really does make a difference. That's easier said than done however, as their impact is usually non-financial and often intangible.

This presentation, based on research with a variety of not-for-profit organisations and case studies of work which Z/Yen has carried out across the sector, identifies four key themes for not-for-profit organisations seeking to evaluate their work and demonstrate its impact. These are:

- Aspiration versus imposition - balancing the requirements of funders and partners with an internal desire to learn and improve performance
- Outcomes not outputs - understanding how to identify the outcomes from the work being done, rather than the volume of activity and outputs
- Gathering Evidence - controlling the process of obtaining evidence so it does not become overwhelming
- The Four Cs - communication, contribution, consensus, commitment.

Through a discussion of these themes and analysis of work being done by organisations who are finding practical ways to provide evidence of their worth (including The Children's Society, the Marine Stewardship Council), this presentation will help attendees understand current and emerging best practice in evaluation of their worth. The presentation concludes by examining the potential of portfolio analysis and programme management as practical approaches that can help evaluate and manage ongoing work, by placing evaluation and learning at the heart of what organisations do.

References
Ian Harris, Michael Mainelli and Mary O'Callaghan, "Evidence of Worth in Not-for-Profit Sector Organisations", Journal of Strategic Change, Volume 11, Number 8, pages 399-410, John Wiley & Sons (December 2002).
Empirical analysis of data pertaining to the effects of changes in marijuana laws is rare.1 While there has been some research on the effects of changes in marijuana laws in Europe and Australia, there are indeed very few such studies in the United States.2,3 Thus, there is a clear need for empirical research to inform the decision-making process in this area of public policy. In the case of medical marijuana laws, critics have argued that these send a message to the public that use of the drug is acceptable and may even have beneficial health affects. If this hypothesis is correct, then one would anticipate greater marijuana use and attendant problems to follow the introduction of medical marijuana laws. We test this hypothesis by examining the effects of the introduction of medical marijuana laws in California, Oregon and Colorado on arrestees using data from the US Department of Justice’s Arrestee Drug Abuse Monitoring Program. Specifically we compare longitudinal trends in positive drug tests among arrestees in five cities in these states (Long Beach, San Francisco, San Diego, Portland and Denver) with 16 comparison cities in states that have not introduced medical marijuana laws. In the analysis we use both regression and time-series models. The presentation will discuss the findings of our analysis of these trends along with a discussion of the challenges of evaluating the effects of changes in laws pertaining to marijuana use.


Developing methodologies to evaluate anticipated consequences of program action

Dr Jonathan Morell  
Senior Policy Analyst  
NewVectors, USA  
Email: jamorell@jamorell.com

This round table is designed to further evaluators' thinking about how their methodologies might better cope with unanticipated consequences of a program’s actions. The intellectual foundation for the discussion is an article that appeared in the December 2005 issue of the American Journal of Evaluation titled: “Why are there unintended consequences of program action, and what are the implications for doing evaluation?”

While programs are in a continual state of evolution, many (but not all) evaluation methodologies require a degree of stability in both program and evaluation design. When this occurs, “connections” between program and evaluation are lost. One solution is to abandon inflexible evaluation methods. Doing so, however, removes many useful tools from the evaluators' toolkit.

The intent of this round table is threefold. First, to present the ideas in the article for critique and development. Second, to find real cases and explore how evaluation might have better served its stakeholders. Third, to begin to coalesce an invisible college of evaluators with an interest in developing thinking about the evaluation of unanticipated program impacts.
Evaluation and Exclusion from the Public Arena: the Case of the British Deaf Community

Dr Sarah C.E. Batterbury
Centre for Deaf Studies
University of Bristol, UK
Email: s.batterbury@bristol.ac.uk

The public arena is an important forum for exchange of information and ideas as well as a battleground for decisions about public policies. It includes widely differing perspectives determined by different traditions, cultures and discourses. Access to avenues for shaping public policy requires a degree of compliance with the prevailing hegemonic discourse of the policy makers: this legitimises some voices and at the same time marginalizes others. The paper discusses how the prevailing policy discourse impacts on the kinds of evaluation likely to be commissioned as well as which evaluation findings come to be used in formulation of new policies. This can result in the exclusion of some evaluation findings and the views of stakeholders from the policy system.

The paper focuses on the ways in which the discourse trap perpetuates the marginalisation of the views of the British Deaf community. This illustrates the way in which their concerns and wishes are ignored by a policy discourse that inaccurately labels the community as ‘disabled’. The Deaf community sees itself rather as a linguistic minority, an epistemology wholly at odds with prevailing policy discourse (Ladd 2003). Existing evaluations and reviews are drawn on to show that where the policy discourse is reproduced in evaluation findings this is at considerable cost to the Deaf community and to the quality and utility of the evaluation findings produced (Nind, Wearmouth, Collins, Hall, Rix, Sheehy 2004). Where evaluators accept the Deaf view, the impact of their findings is minimized (Van der Wilt, & Reuzel, 1998).

The examples demonstrate the difficulty encountered for the Deaf view to find an audience in policy circles in the face of a hegemonic discourse that persists in categorizing Deaf people as disabled in spite of the Deaf community’s assertion of linguistic, cultural and historical patrimony. It shows how the Deaf voice struggles to achieve recognition in evaluation and ultimately in policy decision-making. This illustrates the nature of the challenge for evaluation research in finding a way around the policy discourse trap.

1 This paper reports on a chapter forthcoming in the INTEVAL book Breul J.D., R. Boyle, and P. Dahler-Larsen (eds.) (in press), Open to the Public: Evaluation, Media and Citizens, Somerset NJ: Transactions Publishers

References


Sustainability: A Value in Evaluation?

Daniela Schroeter
Western Michigan University
WMU Evaluation Center, USA
Email: daniela.schroeter@wmich.edu

Sustainability is a fuzzy though ubiquitous concept. As evident in the emergence of varying conceptualizations of sustainability in evaluation-specific publications, the evaluation of sustainability receives growing attention in the evaluation community. However, neither the concept nor mechanisms for evaluating ‘sustainability’ have been clearly established or are agreed upon. On the one hand sustainability is discussed rather differently in specific disciplines and related contexts. On the other hand, it carries divergent meanings in different evaluation domains. Is sustainability a means to an end (i.e., comprised of mechanisms that enable survival of the evaluand or parts thereof), an end in its own right (i.e., the ability to function after initial support mechanisms are withdrawn), or both?

This question suggests at least two distinct perceptions, in essence connotations that call for the differentiation of terminology, such as sustainability versus sustainment. Building linguistically on this differentiation, I argue that sustainability comprises procedural elements (such as social, financial, or political capacities among others) and spatial and temporal characteristics of evaluands that facilitate their sustainment, which - on the other hand - can be described as a concrete, observable outcome or state of being.

This postulation, however, also suggests that sustainability traverses all key components of evaluation, thus raising the question, if the concept ‘sustainability’ should be considered as a ‘value’ dimensions in its own right (cf. Scriven, 2005). A comprehensive literature review across multiple fields and different conceptualizations of and methodological approaches to evaluating ‘sustainability’ will serve to describe sustainability, sustainment, and their evaluation.

A case study exemplifying the evaluation of sustainability in the context of an educational multi-site evaluation in the United States will be used to stimulate discussion and further enlighten concepts and procedures to enhance the evaluation of sustainability.
If 'good' questions cannot be quantified and 'quantifiable' questions are not that good: Encompassing multifunctionality goals in policy evaluation and assessment

Dr Karlheinz Knickel
Senior Researcher
Institute for Rural Development Research at Goethe University Frankfurt, Germany
Email: knickel@em.uni-frankfurt.de

Co author
Melanie Kröge
Deputy Director, Manager
Institute for Rural Development Research at Goethe University Frankfurt, Germany

In the Agenda 2000 it is, for the first time, an explicit goal to secure a «multifunctional, sustainable and competitive agriculture throughout Europe». The increasing importance of multifunctionality goals in support programmes and measures at EU, member states and regional level that is related to this, poses a challenge for the evaluation and assessment of policy. Classical quantitative instruments and tools of policy evaluation appear to have major restrictions in their capacity to deal in an integrated manner with the various dimensions of the multifunctionality of agriculture and rural space. In our contribution we want to critically discuss the appropriateness of relevant conceptual frameworks and evaluation practices for addressing multifunctionality goals.

When discussing the question in how far methods of evaluation and assessment are up to the new requirements that themselves reflect current political and societal trends and changes we will refer to the results of two recent EU funded research projects. Both projects – Multagri and Sustainability A-Test – were dealing with the identification and advancement of appropriate evaluation and assessment methods.

In our contribution we first address the demand side of policy evaluation. We refer to the practical side of the multifunctionality debate and the related sustainable development concerns. This is followed by a critical appraisal of current evaluation practice. We refer to the needs of policy processes and evaluation interfaces, and we discuss research struggling with the needs of a more integrated assessment. In the final section we conclude with some promising lines of research (and thinking), and we formulate concrete recommendations for a more integrative policy evaluation and assessment.

Reference is made to platforms for collective learning and decision-making, the triangulation of qualitative and quantitative approaches, and the need for more consultative, discursive approaches that are supported by hard data. We question the claim for "objectivity" – in the meaning of one reality and verity that can be captured by science – in the solving of complex problems or the assessing of complex policies. We understand integrated assessment as an interdisciplinary process of combining, interpreting and communicating knowledge from various scientific disciplines and/or stakeholders to inform decision-makers on complex societal problems. Our contribution is meant as a constructive input into the ongoing discussion about the advancement of conceptual frameworks and tools towards more integrative approaches.

For further information on the projects see www.multagri.net and www.sustainabilitya-test.net
Evaluation of the effectiveness of the public aid for tourism development: the case of the Campania Region

Dr Federica Bertamino
Senior Economist
Public Investments Evaluation Unit - Campania Region, Italy
Email: f.bertamino@regione.campania.it

Dr Antonella Perillo
Junior Economist
Public Investments Evaluation Unit - Campania Region, Italy
Email: antonellaperillo748@yahoo.it

The present work has the aim to evaluate the effects at local level of the public aid to the enterprises in the sectors of the tourism, of the industry and of the craftsmanship in the Campania Region, that is one of the Objective1 Regions in Italy.

The first footstep of the analysis will involve the data collection and the choice of the indicators that are mostly able to underline the effects of the development public policies at local level. The following phase will concern the building of a matrix in which to bring all the indicators; the matrix will be normalized allowing the comparisons among the variables.

The elaboration of the picked data and their interpretation will be done using statistic techniques like the Analysis in Principal Components and the Cluster Analysis. The aim of the Analysis in Principal Components is to reduce the set of the indicators in a way to represent the most part of the relevant information for the territorial analysis. The Cluster Analysis will be used to apply the indicators to the areas of influence and to divide the statistic unities in homogeneous areas. The V-test will help to determine how much the average of the observed variables for the statistic unities that compose the group is differentiated in meaningful way from the general average.

The evaluation of the Structural Funds aid to the enterprises through the use of these statistic techniques can be of notable help for the ex ante, in itinere and ex post evaluation of public policies.
Principles and Purposes of European Union Cohesion Policy Evaluation

Dr Sarah C.E. Batterbury
Centre for Deaf Studies
University of Bristol, UK
Email: s.batterbury@bristol.ac.uk

This paper presents a critical assessment of EU Cohesion Policy focusing on the current regulatory framework and the difficulties this poses for achieving rigorous and useful evaluation outputs. The paper argues that the evaluation framework for EU cohesion policy is limited to three core purposes: accountability, improved planning, and enhanced quality and performance but that it would benefit from widening this to include other functions. The decentralisation of evaluation to the Member States means the evaluation of cohesion policy relies on the presence of a pre-existing evaluation culture and skills base in the regions. Further, obstacles to effective evaluation arise from the lack of data comparability, rigidity of time-scales, and a focus on performance approaches.

Cohesion Policy is a highly complex policy arena that is difficult to evaluate effectively within the fairly constrained budgets and time frames available. The practicalities of implementing the current tightly regulated, time-defined, decentralised evaluation framework have paradoxically limited the innovative potential for evaluation. Ideally the evaluation function for Cohesion policy needs to mature beyond the confines for assessing impact or even operating as a management tool to become a source of evidence to inform better programme and policy decision-making.

In conclusion the paper advocates that policy and programmatic learning needs to be at the core of the evaluation framework. This would ensure that evaluation endeavours contribute to the emergence of a coherent and useful explanatory evidence base likely to benefit future policy design.

1 This paper was published in a special edition of Regional Studies (2006), edited by John Bachtler and Colin Wren – April, Vol 40. number 2, pp.179-188
Evaluation and the preparation of structural funds new programming period in France

Nicolas Gillio
Manager
Mazars and Guerard
Email: nicolas.gillio@mazars.fr

In the context of the new programming period for European structural funds, we have implemented evaluation tools for different regions in order to define guidelines and priorities for "employment and regional competitiveness" objective.

These methods were supposed to help in defining a restricted number of priorities and in concentrating public policies on a short number of themes. SWOT analysis, benchmarking and focus groups were among the number of evaluation tools we used in order to meet the goals for prioritising interventions and for building a social and economic diagnostic in partnership with stakeholders. This process also tried to give birth to a close relationship between strategy (stakes and goals) and evaluation outcomes.

The author of this paper has used his own experience of evaluator in different regions to describe relationships between evaluation and the European project as far as the Lisbon agenda is concerned and the role of evaluation in policy-design.
Evaluations to improve regional development policy

Dr Lars Niklasson
 Analyst
 Swedish Agency for Public Management, Sweden
 Email: lars.niklasson@statskontoret.se

Regional development is a field of policy where governments try to join-up resources through various kinds of collaboration. Many countries have opted for horizontal collaboration among regional actors as well as vertical collaboration between regional and central governments. Several unitary states have established a regional level of government to take greater responsibility for policy, but most resources are still distributed by the central governments.

This situation creates a need to design systems of monitoring and evaluation to encourage the integration of resources and an enhanced learning capability. For example, the Swedish government is looking for ways in the budgetary system of setting and monitoring joint performance targets. There is also an ambition that the regional partnerships should commission evaluations which will help them improve policy implementation, i.e. to support regional development (primarily economic growth). To some extent, it is a question about evaluation of and by public policy networks.

The purpose of this paper is to provide an overview of how systems of monitoring and evaluation are designed in the field of regional development policy in a number of countries. The ambition is to compare the Swedish situation to other unitary European states (Denmark, Finland, Norway, France, Italy, the UK and the Netherlands) as well as to some federal states (Germany, Australia, Canada and the US).

The paper will discuss the importance of three background conditions:

- whether regional policy is general for all regions or targets regions with problems (development or redistribution)
- to what extent the relevant policy tools are within the field of regional development or controlled by other policy segments (the extent of fragmentation)
- the techniques used by central governments to set the stage for collaboration with and within the regions, especially how it empowers regional governments and how it designs the funding systems with its budgetary controls

It is assumed that these institutional factors don’t determine the role of monitoring and evaluation, that there are good examples for governments to learn from and that such learning is taking place. I hope to be able to discuss how the paradigms of New Public Management and Governance have struggled to shape the systems of monitoring and evaluation.

The ambition is to make a contribution to the discussion on the use of evaluation in public policy by looking at the new situation of joined-up governments in this particular field. It is also hoped that it will add to the literature on governance by focusing on evaluation as a tool for integration. The ambition is to draw conclusions for Sweden in particular.

The study will be based on literature (such as an overview by the OECD) and study visits to the selected countries. My background is in Political Science and I’ve been teaching on some of these matters at the universities of Uppsala, Sweden and Pittsburgh, USA.
The challenge of formative evaluation in complex policy environments: the case of community strategies in England

Alison Darlow
Principal Researcher
Policy Research Institute
Leeds Metropolitan University, UK
Email: a.darlow@leedsmet.ac.uk

Co authors
Martin Purcell and Sukky Jassi
Leeds Metropolitan University, UK

Formative evaluation presents a number of challenges for the evaluator, in particular managing expectations between the evaluator and the evaluand, the ability to innovate and present clear-cut, workable and timely policy solutions as well as good communications between all parties. These challenges are particularly acute when undertaking evaluations within complex policy environments that are experiencing rapid change and where there are multiple other policy innovations occurring simultaneously.

These issues are explored through a major evaluation being undertaken within England on behalf of the Office of the Deputy Prime Minister. Community strategies are one of the key policy interventions which form part of the wider modernisation of local government within England. Community Strategies are in themselves a complex undertaking; whilst responsibility for their preparation lies with local government, they are designed to draw together key partner organisations (the Local Strategic Partnership) at the local level to work towards ‘promoting or improving the economic, social and environmental well-being of their areas’.

The original remit of the evaluation was to focus upon 18 broad issues and questions, reflecting the complexity of this area, using a mixed method approach to produce a range of outputs for policy clients at timely intervals to inform key policy decisions. In reality the complexity of the policy environment, the rapidly changing nature of interventions, and the need to link the various phases of the evaluation with decision making timescales and cycles has presented a considerable challenge. Key lessons are identified for undertakings of this type drawing on the existing evaluation literature particularly focusing on the improvement agenda.
Evaluation and democratic accountability

Ass.Professor Anders Hanberger
Umeå Centre for Evaluation Research
Umeå University, Sweden
Email: anders.hanberger@ucer.umu.se

The purpose of this paper is to discuss key issues concerning evaluation and democratic accountability. The paper deals with the direct and indirect ways that citizens can hold those in power to account and the implications of various forms of evaluation or public review for accountability. First, the concept democratic accountability is unpacked in order to single out various forms of accountability on the one hand, and the established democratic order and democratic values, on the other. Second, the paper discusses the direct and indirect ways that citizens can hold those in power accountable in relation to different forms of democratic governance. Direct ways refer to elections, referenda and evaluation made by citizens or citizens’ organizations themselves whereas indirect ways refer to authorized and self-authorized evaluations undertaken by local auditors, state inspectors and media. Third, the implications of five public reviewers’ evaluation for democratic accountability, in four Swedish municipalities, are discussed.

One conclusion is that the indirect ways of undertaking evaluations for democratic accountability are not sufficient because some of these ways have a limited value for citizens in coming to judgment concerning accountability. Furthermore, they are not quite compatible with participatory and discursive democracy. Another conclusion is that all public review for accountability does not have the same implications for public policy and democracy. If evaluation for accountability is viewed as a supportive tool for democratic governance and public policy it has to be adjusted to new forms of governance and policy making.
Internal versus external evalautor roles: benefits and challenges

Dr Mia Luluquisen
Principal Consultant
Mia Luluquisen and Associates, U.S.A.
Email: esminia@aol.com

Since evaluators can be either employed or contracted for services in various governmental agencies and private institutions, there are considerations for conducting the highest quality evaluations for these two types of evaluators. For example, “internal evaluators” employed by an institution can benefit from having access to program managers and staff, and thus tailor an evaluation specific to the needs of that institution. On the other hand, “external evaluators” hired from outside the institution can have the advantage of conducting objective evaluations that are not directly influenced by the organization’s staff or stakeholders. All evaluators, however, face the challenge of disseminating evaluation findings to appropriate stakeholders for the purpose of program improvement and course correction.

This paper will thus explore the benefits and challenges of conducting evaluations internally versus externally and the related elements for ensuring appropriate evaluations in different situations.
Do we attend to the social relations among participants in a program being evaluated, and if so, why and how do we relate to program participants and other stakeholders in the evaluation setting? These are the questions I will discuss in a paper on the relational dimensions of evaluation. With respect to the social relations among program people and other stakeholders in the evaluation setting, I will draw on evaluative traditions and histories, which assume that social interactions are important because they are integral to and partly constitutive of the quality and effectiveness of the program being evaluated. With respect to the social relations between the evaluator and various stakeholders, these are most importantly the kinds of relationships the evaluator establishes. These relationships are important because they are partly constitutive of the evaluative knowledge generated and because relations communicate particular values and norms. Several positions will be distinguished to identify kinds of relationships between the evaluator and others in the evaluation setting ranging from approaches in which social relations are irrelevant to or sources of error in evaluative work to approaches that aim to manage and change relations or in which social relations are intrinsically valued.
An evaluation of the role of stakeholder engagement in research

Dr David Harrison
Evaluation Researcher
Learning and Skills Network (LSN)
Email: dharrison@lsneducation.org.uk

Dr Liz Walker
Evaluation Research Manager
Learning and Skills Network (LSN)

This paper reports on an evaluation of a sample of public sector (education) research projects that considers specifically the role of stakeholder involvement / engagement in research. The paper draws on literature from a number of disciplines (e.g. education, professional development; social care) to identify some consistent messages that reinforce the importance of stakeholder engagement, namely:

• research and practice discourses and forms of knowledge are likely to remain separate unless pro-actively brought together
• this is likely to reduce the impact of research - i.e. its capacity to contribute to improvement
• one critical aspect of the bringing together of discourses is the engagement of relevant stakeholders over time.

The evaluation identifies the roles of stakeholders (e.g. funders, commissioners, policy makers, practitioners, beneficiaries) in 9 selected research programmes, and considers the nature and degree of their involvement, and their perceptions about potential research impact. The findings are then considered in the context of the issues identified above, and with a view to developing further understanding of the role of stakeholder engagement in contributing to research impact.

The evaluation identifies critical success factors for effective stakeholder engagement, and reports on a number of key lessons emerging about the value and complexities of stakeholder engagement. The process of involving stakeholders early and efficiently is complex and time consuming and requires early planning, focus and engagement with the research commissioner and other stakeholders. If effective engagement is achieved, stakeholders are a significant source of expertise contributing to the design, implementation and outcomes of the study concerned, as well as contributing to bringing about the intended impact on policy and practice.
Quality, evaluation; policy, theory and practice in the education sector

Aidan Kenny
Project Manager Skills Research Initiative
Dit Institute of Technology, Ireland
Email: aidan.kenny@dit.ie

Dr Gerry McNamara
Head of School
The School of Education
Dublin City University, Ireland

Quality assurance procedures and evaluation processes have become dominant mainstream trends within the education sector(s) over the last decade. Theory and practice relating to quality assurance and evaluation have been incrementally embedded into policy at, 'macro' European level, 'mesco' national level and 'micro' institute/school level. This paper sets out to contextualise the rise of contemporary quality assurance and evaluation as evident in European Union and OECD educational policy documents, Irish national policy, and the academic literature.

The mode of inquiry is informed by Guba and Lincoln's 4th generation evaluation and what they term the constructivist paradigm and Yin's case study methodology. The inquiry method utilised comprises of analysis of policy documents, an academic literature review and specific examples from two case studies; one case study focuses on second level school evaluation while the other case study describes a quality review of a higher education institute.

The two authors worked in a collaborative fashion. By combining our experiences from the separate case studies and by analysis of relevant documentation we endeavour to forge a 'triangulation' of multi-dimensional perceived evidence into a framework which we suggest gives credibility and rigor to our inquiry and the conclusions outlined. Our conclusion or more appropriately recommendations are set out in a useful check list format in other to give the reader a practical set of tools to utilise, either as they are presented or customised to suit their own needs. These recommendations are based on both the authors' practical experience of quality review and evaluation in the both second and third level educational sectors and our theoretical understanding gained from engaging with both the literature and the dynamic social processes we actively witnessed. In essence we present a form of theory and practice in action or 'praxis'.
Changes in the wake of a self-evaluation in four schools: A longitudinal study

Dr Sigurlina Davidsdottir
Associate Professor
Faculty of Social Sciences
University of Iceland, Iceland
Email: linadav@hi.is

Penelope Lisi
Connecticut Central State University, USA

Two university professors, one Icelandic and the other one from U.S. assisted four Icelandic schools doing self-evaluations. They used an empowerment approach and deliberative democratic evaluations. Simultaneously, they mapped school changes during this effort. The present study is a case study of the four schools for the first four years.

The findings indicated that teacher attitudes toward the tested domains (school culture; school planning (including sources, goals and monitoring of goal achievement); professional growth; systematic data collection; and school decision-making) improved for every domain from year to year during the self-evaluation effort, using the methods above. Method fidelity differed among the schools. Those schools that adhered best to the suggested methods improved the most between years. The schools who used those methods were farthest along in institutionalizing self-evaluations. Extended-term mixed-methods were used to map the changes and compare between schools. Differing results were to a large extent explained by differing method fidelity.

The method that was used here seems to predict positive results for the schools. Thus it seems that an evaluation of this type could serve to improve and inform practice in schools.
Learning about Networked Learning Communities

Dr Sonia Ben Jaafar
Research Associate
Aporia Consulting Ltd, Canada
Email: sjjaafar@oise.utoronto.ca

Co authors
Lorna Earl, Steven Katz, Susan Elgie, and Leanne Foster

This paper reports on the evaluation of the Networked Learning Communities (NLC) programme in England. The programme was designed as a research and development initiative to inform the broader policy audience about what works when schools intentionally work together. We conducted a three-phase theory-based evaluation designed to explicate promising features and processes of networked learning, and to investigate the importance of these features in enhancing teacher practice and pupil success. Phase 1 engaged international experts in sharing existing knowledge in the field to document what is known and what deserves continued attention and consolidation. Phase 2 focused on drawing on current knowledge to construct a conceptual model of key features of networked learning communities and gathering data from a small number of schools engaged in NLCs. Phase 3 empirically examined the integrity of these key features, the relationships that exist among them and their relative impact on knowledge creation, changes in practice and pupil attainment.

Education is only one of many public domains whose policies are promoting networked learning communities as a key component of creating systemic improvement through collaborative capacity-building endeavors. The findings of the evaluation offer insight on the operational influence of key features of networked learning on professional knowledge building, beliefs, practices, and ultimately achievement outcomes. The consequences of this evaluation were tightly linked to its purposes, which were to 1) build participant internal evaluation capacity, 2) describe the conceptual theory of action for the programme, 3) examine the impact of networked learning communities on learning at all levels, and finally 4) to support policymakers in making evidence-informed decisions regarding future policy and program decisions. The innovative design builds on current theories and practices for evaluation that emphasize the significance of making the results useful for informing policy decisions.
Evaluation of mainstreaming social inclusion in European Union member states

Izabela Litewska
EU Project Researcher
Combat Poverty Agency, Ireland
Email: litewskai@cpa.ie

The European Council held in Lisbon in 2000 agreed that to tackle poverty and social exclusion successfully, all EU Member States would adopt a poverty perspective in their policy-making, thus placing mainstreaming social inclusion at the centre of the European Social Model.

An evaluation of how social inclusion is mainstreamed into public policies within the EU Member States has not been undertaken until recently due to the lack of an evaluation tool to measure mainstreaming social inclusion (Combat Poverty Agency, 2005). There are apparent difficulties in measuring social policies, as their impact may not be evident in the short term and could only bring change over a longer timeframe, outside the scope of the evaluation process. Consequently, it would be necessary to have a sequence of evaluation studies, over time, in order to establish a clear picture of the policy effectiveness.

This paper presents the outcomes of the research project supported by funding from the European Commission (2003-2006) and co-ordinated by the Combat Poverty Agency, Ireland. It examines how social inclusion is being mainstreamed into the public policies of a number of countries. In particular, it focuses on methodology that was developed to measure the level of mainstreaming of social inclusion in policy-making and the approach that has been currently applied under a wider European evaluation study. It concludes with the presentation of preliminary findings arising from the research, specifically concentrating on the following objectives:

- to provide a benchmark for the measurement of the eradication of poverty and social exclusion;
- to develop and to test a methodology and tools for assessing mainstreaming social inclusion;
- to evaluate the commitment to the eradication of poverty and social exclusion in public policies at each level of governance;
- to compare different structures of governance in the participating countries, with regard to the design and implementation of policies;
- to establish what the understanding of the definition of mainstreaming social inclusion is across the participating countries.
Making the connections: evaluating the EU Community Action Programme to Combat Social Exclusion

Frances Abraham
Principal Consultant and Senior Researcher
The Tavistock Institute
Email: f.abraham@tavinstitute.org

The EU Community Action Programme to Combat Social Exclusion is part of the ‘soft policy’ interventions of the Open Method of Coordination. The Programme included Peer Review involving national administrative officers; Transnational Exchange Projects involving local and regional actors as well as sometimes national policy actors; the development of statistical data through the EU Survey on Income and Living Conditions (SILC); Round Table events; European Networks; Studies; and a number of other actions.

The Programme can be seen as a large transnational learning system on the subject of social exclusion, which brings into interaction EU and MS National, Regional and Local Administration, NGOs and Civil Society in learning from each other around social exclusion issues. Critical points of shared learning ensued in relation to cross-thematic issues, often involving particular target groups which may be addressed by different agencies and agencies at different levels. National contexts for policy making and practice adoption varied widely as shown in specific MS cases presented, in the strategic European context of the OMC, including National Action Plans for social exclusion. These were not always structured, intended or followed through. Mobilising the support of national administration was often a key to realising the learning of local actors involved in Transnational Exchange Projects for example, but was not always considered at the outset.

This paper gives an overview of the types of different learning and the ways in which learning occurred, using concepts and frameworks from learning theory and organisational learning.
Evaluation practices and primary school. A research on how teachers use evaluation at school and students perceive it

Professor Antonella Nuzzaci
Universita della Valle d’Aosta, Italy
Email: a.nuzzaci@univda.it

It is not possible to keep didactics separated from evaluation, because these two dimensions represent two sides of the same coin; nevertheless the different functions of the evaluation (diagnostic, instructive, supplementary etc.) often remain arbitrary expressions, not connected with reliable procedures and valid instruments.

The research – conducted in three Italian cities (Rome, Siena and Aosta) – studied the use of evaluation carried out by primary school teachers, and how students perceive it. In particular, the research has explored the evaluation methods used in terms of teachers’ times, resources, instruments and descriptions.
Peer Review - Peer Evaluation?

Sandra Speer
Univation Institute for Evaluation, Germany/Hungary
Email: Sandra.speer@univation.org, sandra_speer@t-online.de

Peer Reviews are in some policy fields widely spread in others relatively new. Various forms of Peer Review have been developed and are in practice. The objects of a Peer Review can be institutions, programs and systems as well as policies. Sometimes “Peer Review” and “Peer Evaluation” are used as synonyms. But are really all Peer Review also evaluations?

Peer Reviews will be distinguished concerning different levels of analysis and underlying concepts. Examples will be drawn from the field of vocational education and training (VET).

1. Peer reviews of teachers in classrooms may evaluate their respective teaching quality. Here an example from the Leonardo da Vinci Project “TEVAL: Evaluation Model for Teaching and Training Practice Competencies” will be drawn.¹

2. Peer Reviews like in the framework of the Leonardo da Vinci Project “initial vocational education and training” may focus on the VET organizations and institutions.²

3. Peer Reviews within the European Open Method of Coordination (OMC) look at national policies and programs, e.g. labor market training. Here several OMC practices within different policy fields exist.

The further focus of the presentation will be: In how far are the different Peer Review approaches interlinked to other evaluation approaches or systems? Most of the Peer Reviews need self-evaluations as a prerequisite other are e.g. established in the context of benchmarking processes.

Finally: Which evaluation standards are relevant for the different kinds of peer review? For some Peer Review approaches different evaluation standard sets should be taken into consideration: Student Evaluation Standards, Personnel Evaluation Standards and of course the Program Evaluation Standards, which will be shown for the different types of Peer Review.

¹ Leonardo da Vinci Project PT/04/B/F/PP-159050
² Leonardo da Vinci Project AT/04/C/F/TH-82000
Industrial Policy Evaluation in Developing Countries. A methodological framework to assess the effect of Export Processing Zones: a preliminary study of Egypt

Dr Elisa Barbieri
Junior Researcher
University of Ferrara-Faculty of Economics, Italy
Email: e.barbieri@economia.unife.it

The aim of this paper is to extend some of the arguments developed by the evaluation literature, in industrialised countries, to developing contexts. The origin of evaluation as an institutional practice is commonly attributed to the American experience, however in the last two decades the debate over policy evaluation has also spread throughout Europe, and the need for a methodical assessment of public policy is felt in European countries as well. On the other hand, public authorities in developing countries are not yet promoting the diffusion of an evaluation culture and public policy intervention in most of these countries is often independent from both ex-ante assessments of expected results and ex-post evaluations of its impact.

The specific objective of this paper is to define a methodological framework to evaluate one particular tool of industrial development policy, export processing zones (EPZs), that has been increasingly used by developing countries in order to promote industrial development through private investment and exports.

The paper focuses on issues of impact evaluation and counterfactual estimation. As the prime objective of EPZs is to make a targeted area of a country particularly attractive by creating a cost advantage for firms, the paper suggests a methodological framework to evaluate the effects of EPZs on firms’ production costs. Firms inside an EPZ should display lower production costs than they would experience outside the EPZ and the paper suggests that a staffy should analyse firms outside the EPZ (but in the same country) as a "control group" that approximates the counterfactual situation of firms inside the EPZ.

In order to verify the feasibility of a similar analysis on real data, the case study of Egypt was chosen and eleven firms were interviewed. Some of them were inside an EPZ and others were not. The limited number of observations does not allow to obtain statistically significant results concerning the effect of EPZs. However, the results offer some suggestive and interesting insights on the functioning of EPZs. Our findings deal with the relevant variables to consider in evaluating EPZs and on the capacity of Egyptian EPZs to attract investment. Furthermore they confirm the feasibility of expanding the study at reasonable costs on bigger samples by using the same methodology.
An evaluation of Japan's junior secondary school construction project in Indonesia

Atsuko Toi
Tokyo Institute of Technology, Russia
Email: toiatsuko2@yahoo.co.jp

Developing countries have long been troubled by unavoidable choice between providing broad access to education and developing high-quality school. However, a review of the empirical evidence from the Junior Secondary School Construction Project in Indonesia, which was the first and largest investment by Japan in the area of basic education in Indonesia, suggests that this is a bad way in thinking about human resource development. This article points out that the legitimacy of the school construction project should not be discussed from the view point of the quality-quantity tradeoff; rather, it should be discussed if quantity improvement by the project is accompanied with quality improvement.

The article first examines the quantitative aspect of the project. The results show that the project appears to have made a significant contribution to junior secondary school enrolment ratio, even during the economically difficult period in Indonesia.

Secondly, it evaluates the qualitative aspect of the project by using the covariance structure analysis on the relationship between educational environment and outcome.

The main finding of this section is that newly constructed schools tend to have weakness, especially in teachers and finance, which is found the obstacles for the sustainable development of the project. The results of analysis indicates that, for the first stage of the school development, its physical environment tends to have positive impact on students’ performance, while teacher qualification will become a relatively important factor for improvement of students’ performance later at the stage in Indonesia.

Thirdly, based on the findings in the previous sections, the paper presents a set of recommendation for the future school construction projects, emphasizing the importance of changing the concept from "providing new buildings" to "establishing new schools with beneficiaries". In order to achieve the desired impact on education outcomes, the future projects should go beyond simply improving infrastructure, and give more attention to the "software" of education as an integrated intervention.
Methodology of project appraisal for small infrastructures: a Croatian case study

Silvia Vignetti
CSIL-Centro Studi Industria Leggera, Italy
E-mail: vignetti@sagittea.com

Alessandro Valenza
CSIL-Centro Studi Industria Leggera, Italy

In Zadar and Sibenik–Knin counties, in Croatia, a Sustainability Development Project funded by CARDS is active since 2003 for the war affected areas. The project aims to support the two administrations in managing local development programs through CARDS and to prepare for the pre-accession Funds. The assistance consists in training in EU procedure, preparing grant scheme, updating the Regional Operational Program, implementing an infrastructure project pipeline. We will focus on the last activity and we describe the experience developed in this context through a methodology commonly used for the management and evaluation of initiatives promoted by Structural Funds. The challenge of this experience is represented by the reference context, where either Public Administration than local actors started only recently to cope with the EU program procedures and principles, and by the chosen operative methods, through a cooperation of International experts and professional evaluators with local personnel.

The methodology applied for the evaluation was focused on a cost-benefit analysis. Given the financial dimension of the concerned projects (from a few hundred thousands EUR to a maximum of one million EUR) a simplified yet rigorous version of the standard methodology has been applied. For the data collection we dealt with the contribution of local actors, either the personnel of the company appointed for the management of this program than the Public Administration and the projects promoters. Surveys, personal interviews through well-structured questionnaires, printouts’ desk analysis and presentation of the main results were the activities of the involved personnel.

The application of a standard and rigorous methodology allows a certain possibility of comparison among the projects and makes it easy to horizontally read the interventions’ performance. The paper presents the main results obtained out of a total of around thirty projects, and deals with some of the specific issues encountered, from tariff choice to the level of Public Fund through the demand forecast.
Strengthening the UN’s evaluation function in the CEE/CIS region

Muhammad Usman Akram
Regional Evaluation Advisor
UNDP Regional Centre for Europe and Commonwealth of Independent States (CIS)

Marco Segone
Regional Advisor, Monitoring and Evaluation
UNICEF Regional Office for CEE/CIS & BS
Email: msegone@unicef.org

UNDP and UNICEF aim to explore ways to complement regional efforts for improved, transparent, effective and results-oriented evaluation approaches and practices for greater development effectiveness. Followed by UNICEF presentation on New Trends in Development Effectiveness in Eastern Europe and the CIS countries, this presentation will aim to reflect on UNDP shift to development evaluation globally as well as regionally, in wake of the changing development aid scenario and the wider UN Reform agenda. UNDP endeavours to internalize the UN Evaluation Norms and Standards, and their subsequent contextualization to UNDP scope of work nationally, regionally and globally.

The presentation attempts to share practical measures that UNDP proposes to strengthen evaluation function that contributes to meaningful participation of stakeholders for greater ownership and accountability, while enhancing the quality, effectiveness and usefulness of evaluation function.

In terms of structure, the presentation will look at changing priorities and trends in development evolution; shift from conventional aid evaluation to development evaluation and what it entails in practice; the present approach and measures being undertaken to address evaluation for improved development effectiveness. Specific references will be made to global Evaluation Network; introduction of quality standards for improved evaluation and plans to enhance the evaluability of UNDP supported development programmes in Europe and the CIS.
Building capacity, lessons from the audit world

David Goldsworthy
Manager, International Technical Co-operation
National Audit Office, UK
Email: david.goldsworthy@nao.gsi.gov.uk

Over the past five years, the UK National Audit Office has devoted substantial efforts to helping strengthen external public audit in developing and transition countries. While early approaches tended to focus on building technical skills through running training courses, supervising pilot audits and arranging study tours and placements, the trend now is to focus more widely on institutional change. There is a growing recognition of the importance of seeking to modernise the organisational environment in which the audits are being carried out, for example by helping build management, human resource development and IT systems and skills and by working with external agents to fuel the demand for high quality audit reports.

The lessons emerging from this work may have direct relevance to attempts by the evaluation community to help build evaluative capability in similar countries.
Evaluation in the European Commission: a comprehensive evaluation framework in the face of complexity

Svend Jakobsen
European Commission
DG Budget

Christel Schröder
European Commission
DG Budget

The European Commission has a comprehensive evaluation framework which embraces expenditure programmes, regulatory measures and thematic issues. The focus is on the activities targeted towards citizens, business or other groups outside the Europeans institutions. Evaluations are used both for policy development and for assessing performance once an activity is implemented. The evaluation system spans all the EU’s policy areas. Moreover, it embraces several different types of management ranging from direct management by the Commission to activities implemented in collaboration with other bodies or with the Member States.

The evaluation framework of the Commission responds to this complexity by combining decentralised responsibility for the evaluation of individual policy areas (each Directorate General must ensure regular evaluation of its activities and follow-up on the results) with central-level coordination concerning capacity building, cross-cutting reporting and strategic-level use of evaluation findings.

The evaluation system is highly institutionalised. The main rules (e.g., the obligation to evaluate, evaluation capacity, evaluation standards) are set out in various regulations, for example in the Financial Regulation. Moreover, evaluation planning and the use of valuation results are integral parts of the Commission’s annual planning cycle and they are also reported on to the legislative budgetary authorities of the European Union.

Most evaluation studies in the Commission are outsourced and the presentation will provide insights into the volume and nature and potential of this market.

External studies recognise the European Commission as an institution where an evaluation is well developed, while its evaluation framework continues to adapt to needs. Evaluations and follow-up works adequately at the level of individual instruments, in particular for expenditure programmes. However, there is room for improving the strategic level use of evaluation, i.e., for informing about performance regarding broad strategic objectives and for the allocation of resources across policies.
Joint Evaluation: Experiences from the Tsunami and other Emergencies

John Mitchell
Head, Active Learning Network on Accountability and Performance in Humanitarian Action (ALNAP) (Chair), and member of the Core Management Group of the TEC

Niels Dabelstein
Head, Evaluation Secretariat, Danida & Ministry of Foreign Affairs, Denmark, and member of the Core Management Group of the TEC

Rachel Houghton
Coordinator, Tsunami Evaluation Coalition (TEC) / ALNAP

Pauline Wilson
Emergency Capacity Building (ECB) Project of the Inter-agency Working Group

As the value of joint evaluation becomes increasingly recognised, this panel will present experiences from two unique joint evaluation initiatives in the humanitarian sector: the Tsunami Evaluation Coalition (TEC) and the Interagency Working Group's Emergency Capacity Building (ECB) project.

The Tsunami Evaluation Coalition (TEC) was established in the wake of the December 2004 tsunami. It is a multi-agency initiative involving over 40 humanitarian and development cooperation agencies, with representatives from UN agencies, multi- and bi-lateral donors, NGOs, the Red Cross Movement, and a number of research groups. The TEC represents the broadest evaluation of a humanitarian response since the Joint Evaluation of Emergency Assistance to Rwanda in the mid-1990s. Its primary aim is to improve the quality of humanitarian action by learning lessons from the international response to the tsunami, and turning these lessons into action. An important secondary aim is to test the TEC as a model for future joint collaborative evaluation.

The Interagency Working Group is a 7-agency consortium comprising CARE International, CRS, IRC, Mercy Corps, Oxfam-GB, Save the Children US and World Vision International. Under their Emergency Capacity Building (ECB) project they have supported four joint evaluations with the aim of building capacity and improving the quality of their response.

Both initiatives offer an important opportunity to learn from joint evaluation processes. In order to achieve this the TEC has already facilitated a number of learning reviews. Moreover, TEC learning is feeding into a wider humanitarian initiative (facilitated by the Active Learning Network on Accountability and Performance in humanitarian action) that is bringing together a much wider array of actors, including the IWG members, to reflect on their experiences with joint evaluation and to contribute to the developing literature on this topic. Work is focusing on how joint evaluations work in practice, what is achieved or lost through them, and how they can contribute to institutional learning and change. While the panel will focus on the TEC, it will also share learning from the IWG experience.

After the presentations, questions from the floor will be managed by the Chair in the form of a group learning exercise, whereby workshop participants will be asked to reflect on the implications of what they have heard for their own practice and to share their experience on how such joint evaluation processes could be improved.
John Mitchell, Head of ALNAP, will open the session as Chair. ALNAP has long had an interest in joint evaluations, and indeed itself came out of the findings of the JEEAR in the mid-1990s. Niels Dabelstein, the most senior member of the panel, was involved in the JEEAR and has thus brought a vast amount of experience to the CMG and the TEC. He will talk about the methodological developments in joint evaluation over the past 10 years and the difficulties and challenges faced by the TEC process, its successes and weaknesses. Rachel Houghton, TEC Coordinator, will briefly present the substantive findings and recommendations from the TEC studies. Finally, Pauline Wilson from the ECB project will compare and contrast their experience of joint evaluation with that of the TEC.
Barking up the wrong tree? Do evaluation methods for simple, stable interventions really work for complex interventions? Towards a framework for evaluating complex interventions

Dr Sanjeev Sridharan
Head, Evaluation Programme
Research Unit in Health, Behaviour and Change
Medical School
University of Edinburgh, UK
Email: sanjeev.sridharan@ed.ac.uk

Using specific evaluations of public health interventions in Scotland, we propose a framework to evaluate complex interventions. This framework consists of the following:

(a) A focus on the networks that link the components of the complex intervention
(b) Explication as an important purpose of evaluation
(c) An emphasis on system dynamics
(d) A focus on exploring stakeholders co-constructions of the intervention dynamically.

The relationship of this framework to recent developments in evaluation including theories of change, theories of influence and realist approaches to evaluation are explored. Key methodological problems that evaluations of complex systems need to address are identified; possible solutions are also explored.
Assessing Impact in large-scale complex programmes

Isabelle Collins
Director
Technopolis Ltd, UK
Email: isabelle.collins@technopolis-group.com

Evaluation of European-funded programmes is increasingly focusing on the question of impact at various levels. This has needed the development of new methodologies to address the issues, which can take into account the increasing breadth and complexity of the programmes while producing coherent results. This paper looks at the lessons learned from implementing a new methodology across a selection of sectors covered by a programme, and how and to what extent the programme’s overall impact can be assessed. It sets out the overall methodology developed as a result of several studies, and draws conclusions on the possible, the potential and the pitfalls.

The challenge was to devise an impact assessment framework relevant to all sections of the programme but which can bring them together at the overall programme level in a coherent manner. Several methodological and pilot studies were carried out, and now the results of these are being brought together in an attempt to implement a first set of impact assessments, which can be aggregated or synthesised at the programme level. Lessons learned range from the methodological in terms of the tools and expectations, to the highly practical in terms of organising and managing a set of assessments independently within a coherent framework. It addresses directly the challenges of choosing between the methodologically pure and the pragmatically possible, of assessing long term impacts in a field of rapidly changing technologies, markets and policies, in an environment where no baselines were set and impact assessment of this nature was not envisaged in any structured way, and where the logics behind the individual priorities are at best not well documented and at worst lost. The methodology also has to be implemented by a range of evaluators from different backgrounds, being clear without being prescriptive and flexible without losing focus.
Evaluation in an indigenous context: New Zealand Maori immersion schools

Frances Salt
National Manager Reporting Services
Education Review Office, New Zealand
Email: frances.salt@ero.govt.nz

Isabell Sinclair Irwin
Education Review Office, New Zealand

The New Zealand Education Review Office (ERO) is responsible for reviewing and reporting on the performance of all New Zealand schools including kura kaupapa Māori (Māori language immersion schools). Some of these kura have been established and operate in accordance with a particular set of principles – those of Te Aho Matua. The evaluation methodology for these Te Aho Matua kura was first implemented by ERO in 2001 and has been recently reviewed during 2006.

This presentation explores the way in which the evaluation methodology has been reviewed through a partnership between ERO and Te Runanganui o Aotearoa (a group charged with promoting Te Aho Matua principles in kura). It also describes and analyses the way in which these particular evaluations are now being carried out in the kura, drawing on the knowledge and expertise of the whānau (the school community) as well as the evaluation experience of ERO.
Collaborative evaluation - when the evidence is not there

Anne Barton
Head of Health and Social Care
M·E·L Research, UK
Email: anne.barton@m-e-l.co.uk

Dr Jill Jesson
M·E·L Research, UK

As with all commissioned research, evaluation for the public sector client poses many challenges. You don’t know until you win the contract and open the files what exactly you are dealing with. We make unfounded assumptions about managers and staff and implicitly expect them to be evidence-based organisations and practitioners.

This paper considers such a dilemma. The task was to evaluate a small community-based intervention in Asian mental health. The purpose was to produce a report which would enable the project, after five years in existence, to be mainstreamed. The evaluation design was simple: first stage documentary review, followed by interviews with key stakeholders and staff.

However, the documentation that we could find was not systematic or robust enough to build a sound picture. The evolution of the project had lead to multiple objectives, which no-one was monitoring. The project idea came from the PCT, the early funding from SRB but later NRF. There was no local needs assessment to justify the project. Over the five years there had been major management shortcomings. We could not honestly match activity against the multiple changing objectives, or quantify any real community outcomes. Community links had been developed, but there was no collaboration with the community mental health team. Yet, we respected the resilience of the staff, who genuinely believed that their work in the community was worthwhile.

The collaborative relationship we had developed with the client meant that although we had to present an honest evaluation report, we also felt a moral obligation to the staff to help them to produce valid evidence. So, in effect we helped them to change.

The roundtable discussion is about the client–evaluator relationship. What do you do when management is the problem? How do evaluators maintain objectivity and at the same time improve public services?
Maximising the value of evaluation and generating buy-in

Karen Miller
Programme Evaluation Officer
The Health Foundation, UK
Email: karen.miller@health.org.uk

The Health Foundation is an independent charity that aims to improve the quality of healthcare being provided in the UK. Over the last 3 years the Foundation has developed its evaluation function, part of which has been the growth of an internal evaluation team. An important aspect of this work has been to encourage key stakeholders to participate in self evaluation activity. This has challenged the evaluation team to make evaluation meaningful and relevant to those concerned to maximise buy-in and collaboration.

Using the Foundation's Leadership Programme as a case study, this presentation will explore the ways in which we have sought to involve different stakeholders in the process of ongoing self evaluation. The Leadership Programme comprises eight schemes which were developed to create a cadre of leaders who could improve the quality and performance of health services in the UK. The stakeholders we have sought to involve in self evaluation include Foundation staff co-ordinating the Programme, Leadership Development Consultants delivering the Programme and Programme participants.

A range of different methods to have been used to make self evaluation meaningful and relevant to these stakeholders, such as developing a self evaluation toolkit for Programme participants, involving Foundation staff in evaluation activities and developing systems to stimulate reflection and feedback from consultants.

The presentation will conclude with key learning points from the process and will invite discussion about further ways in which buy in for self evaluation can be achieved and maintained.
Appreciative inquiry and evaluation – practical guidance for evaluators

David MacCoy
Partner
First Leadership, Canada
Email: david.maccoy@firstleadership.com

Anna Engman
Cathexis Consulting Inc, Canada

Burt Perrin
Burt Perrin and Associates, France

Appreciative Inquiry (AI) has been increasingly used as an approach for conducting evaluations. When it works well, it enhances change through the critical connection between the evaluation and its utilization in the organization. The practicality of an evaluation approach is crucial for enhancing its use among practitioners. Practical, quality guidance on how to use the AI approach in evaluation has been starting to emerge over the last few years, but the field is still developing. How can the AI evaluation methodologies be refined to better meet the challenges the evaluator faces when working on the ground? What are the factors that contribute to it working well? What are the factors that detract? How can evaluations be managed to support the results that can be achieved through successful AI application? A variety of issues will be discussed such as different variations on how the AI approach can be integrated in an evaluation, how the AI approach can help to identify poor performance, skills required when using the AI approach, power balances in the organization being evaluated, the learning processes for those involved etc.

This panel will discuss some of others and their own experiences, providing an overview of successful AI methodologies and looking at ways to manage some of the challenges. David MacCoy will set the stage through a presentation of the principles of AI. Anna Engman will present an overview of some real life examples that demonstrate an AI approach to evaluation, looking at the strengths and challenges. Burt Perrin will serve as a discussant raising questions that point to the need to ensure rigor and consistency with evaluation principles.

The presentation’s purpose is to contribute to the development of general applicable, practical guidance for evaluators who are interested in using or currently are using the AI approach in evaluation.
Theory-linked research approaches for quality evidence about 'what works for whom': three practical examples

Dr Patricia Rogers
Director
CIRCLE (Collaborative Institute for Research, Consulting and Learning in Evaluation),
Royal Melbourne Institute of Technology, Australia
Email: Patricia.Rogers@rmit.edu.au

Exploring Context-Mechanism-Outcome Configurations
Conceptually, Empirically and Pragmatically

Patricia J. Rogers and Bron McDonald
Royal Melbourne Institute of Technology, Australia

The realist notion of Content-Mechanism-Outcome configurations (CMOCs), as described by Pawson and Tilley (1997), is conceptually appealing but can be difficult to investigate empirically. How do we move between conceptual and empirical explorations of CMOCs? How do we identify and label the generative mechanisms that underpin observed patterns between context and outcomes? How do we know how many different configurations to investigate empirically? This paper presents the reconstructed logic of analysis from an evaluation of a dairy extension program. It describes how evidence from a series of mixed method investigations was iteratively developed and used to identify patterns in observed outcomes, and then analyse them in terms of CMOCs. It discusses the ways in which this information was received by evaluation stakeholders, its utility for program decision making - and the diminishing returns from further exploration of CMOCs.

Theory-informed quantitative data analysis in realist program evaluation

Gerald Elsworth
Royal Melbourne Institute of Technology, Australia

This paper picks up a theme in the critical realist literature (in the work of Mingers, Olsen and others) that supports a role for quantitative data and statistical analysis within a mixed-methods framework. As Olsen has argued, the use of statistical analysis should be carefully qualified: inferences must be framed within a realist interpretation of the data generation and underlying social processes, and understood as occurring within open systems. Drawing on an evaluation of the re-development of community correctional services in an Australian State, the paper presents an example of how this analysis approach might be implemented. Econometric and graphical data analyses were conducted on two complementary data sets, informed by an emerging theory about the impact of re-development on the likelihood that magistrates and judges would make decisions that substituted community orders for custodial sentences and the possible mechanisms that were activated. It was hypothesised that redevelopment would result in a decrease in the prisoner-to-offender ratio of monthly offender 'flows'
into the correctional system and that this anticipated impact would be associated, specifically, with certain (drug-related) offence types and particular offenders (those who had previously served a custodial sentence). Analysis was intended to assess whether these theorised regularities were evident in the data. Analysis commenced with a ‘strong’ test of the empirical hypothesis relating re-development to a decrease in the prisoner-to-offender ratio. An econometric ‘event study’ regression model was used. The hypothesis of no association was not supported and the result was argued to provide initial support for the theory. Subsequent examination of possible threats to the validity of the observed association motivated the use of graphical analysis of a second, more disaggregated, data set to examine in detail the kinds of offence types and offenders impacted. Together, the results of these analyses provided strong corroboration of the theory.

Recent Developments in Longitudinal Methods to Understand the Mechanisms of Change over time

Sanjeev Sridharan
University of Edinburgh, Uk

In this paper, we review some recent development in longitudinal methodology to better understand the mechanisms of change over time. Using an example of an evaluation of an alcohol intervention delivered in a fraternity settings, we describe the following methodological concepts that can help identify and understand the mechanisms of change: (a) Developmental Trajectories; (b) Trajectory classes; (c) Programmes as modifiers of risk-outcome linkages; (d) heterogeneity of programme effects. Relationship between our methodological framework and realist approaches to evaluation are discussed.
Ministerial inspectorates: how do they transform or reproduce society?

Dr Friso Coumou
Senior Reseacher
Inspectorate for Work and Income, The Netherlands
Email: fcoumou@iwiweb.nl

Inspectorates have an important role to contribute as an independent evaluator to the improvement of the execution of the policy, the social policy itself and the public debate. This is especially true for ministerial inspectorates. Because these inspectorates act under the responsibility of the minister, they report their judgements directly to the minister. The minister can use the outcomes of the evaluation-study in directing the organizations that execute the social policy. The minister may also decide to change the social policy itself. The executive organizations can use the outcomes of the evaluation-study carried out by the inspectorate for improving professional practices. Ministerial inspectorates also play a role in the public debate. The minister sends the report of the inspectorate to the parliament, where the political debate takes place and is taken up by the media.

Ministerial inspectorates have to deal with several parties with sometimes different interests. This leads to the question in what way these inspectorates are most effective in influencing their environment. What is the best way to contribute simultaneously to improvements in the execution of social policy, social policy itself and the public debate? In this paper I will outline a few perspectives on the role and position of ministerial inspectorates.

Within these perspectives I will describe a few dilemmas and the choices and practices to try to resolve these dilemmas. Amongst the themes are: how is decided on the topic of the evaluation study? Who determines the norms and how? What form is chosen for the final report and how does the inspectorate communicate its findings? I will then move on to the comparison of the perspectives from a societal standpoint. This is done by placing the perspectives within the Transformational Model of Social Activity of Roy Bhaskar (1989).
The National Audit Office of Denmark: report on the government's use of evaluation

Malene Damslund Fiala
Auditor
The National Audit Office of Denmark, Denmark
Email: malene.damslund.fiala@rigsrevisionen.dk

The National Audit Office of Denmark (NAOD) has carried out an audit of the government’s use of evaluation. The NAOD has carried out this particular audit because evaluation is a management tool, which may supply politicians and management with information about the results of government activities. In recent years, it has become increasingly more common to evaluate the results of government activities, and public spending on evaluation reports is significant. Therefore, it is essential that the government can use the results of the evaluation reports, and that the government receives value for money.

The NAOD has examined how the government may strengthen its role as contractor of evaluations and has furthermore focused on the advantages and disadvantages of using evaluation reports as management tools. Consequently, the NAOD has examined whether an evaluation report is a well-defined management tool, whether the quality of the evaluations is satisfactory, whether the evaluations have an effect, and whether the government has practised economy when acquiring evaluation reports.

The NAOD has reviewed 26 evaluation reports within 7 areas of ministries. Some of the conclusions are:

- The quality of the evaluation reports is satisfactory. However, the quality can be improved, e.g. by ensuring a description in the report of advantages and disadvantages of the applied method.
- The evaluations have an effect. This is based on the fact that most of the evaluation reports form the basis of political and administrative decisions. However, the effect can in some respects be enhanced.
- Generally, the government has practised economy when acquiring evaluation reports. Some of the evaluations have, however, caused undue costs for the evaluated part.
When is “summative” evaluation formative?: helping funders foster strategic change

Dr Laura R. Meagher
Senior Partner
Technology Development Group
Email: Laura.Meagher@btinternet.com

Co author
Dr Catherine Lyall
Senior Partner
Information Browser Ltd

When a new scheme aims to bring about strategic change through funding novel activities, evaluation can play a significant role in investigating, capturing, distilling and disseminating the learning that arises. By combining qualitative with quantitative methods, evaluation can not only assess hard and soft impacts of such a scheme; it can also – perhaps more importantly – capture the informed insights and lessons learned of those who participated in the funded “experiments”. If funding agencies take evaluation seriously, such continuous learning as to issues and practical solutions can 1) accelerate evolution of the scheme or subsequent alternatives and 2) enhance effectiveness of future implementation. Thus even seemingly final impact-assessing evaluations can contribute to future change. This role of evaluation underscores key democratic premises: the ability of each individual to make a contribution and the promise of progressing further together than by leaving each individual to face the same problems alone.

Connectivity between evaluation learning and improved capacity to promote change will be illustrated from experiences of the authors with three evaluations:

1. A Scottish grants scheme encouraging multi-faceted centres focussed on emerging research fields
2. A UK Research Council training and development scheme promoting interdisciplinarity
3. A major US foundation’s national scheme to transform land grant universities and their connectivity with stakeholders.

Each funding scheme was itself innovative and intended to stimulate innovation. In both UK evaluations, semi-structured interviews, surveys and focus groups enabled those caught up in the scheme to reflect on core issues, critical inflection points and pragmatic mechanisms. Implementers, often for the first time, became aware of the tacit knowledge they had accumulated as to what works in their particular strategic change processes. Recommendations for funders and for future implementers, shared via evaluations, can help funding schemes to evolve even as future grantees benefit from peer advice.
Concern-based evaluation: examining policy in view of a specific public demand and political commitment

Dr Laurent Mermet
Professor
ENGREF¹, France

Co authors
Dr Raphaël Billé
Freelance consultant, France
Dr Maya Leroy
Assistant Professor, ENGREF, France

This paper presents an innovative framework for evaluation in situations where the main operator of a public policy does not hold leadership in his own policy domain, and has very little control on policy outcomes. In such cases, the “command and control” model does not apply at all, even in its more trendy attenuated versions that attempt to mobilise wider circles of stakeholders and policy operators, in what can be seen as a “joint-command and joint-control” perspective.

Our proposals for reframing policy evaluation are based on our works in the field of environmental policy evaluation, particularly on the French wetland policies, on Integrated Coastal Management policies and guidelines, and on the fulfilment of environmental commitments in development and resource-management projects and programmes in the Senegal valley. They are underpinned by the theory of strategic environmental management developed by our research group.

Environment is archetypal of policy domains where growing complexity in decision-making processes affects the meaning, role and function of evaluation in society.

The case will be made for an evaluation paradigm and methodology we propose to name “concern-based evaluation”: an evaluation based on a public concern or “cause” that is also a reference point for a public policy commitment. With examples from our evaluation studies, we will show how such a framework can be built into evaluation methodologies that (with hard work) deliver answers on decisive questions that were eluded by evaluations based on other frameworks.

In view of current trends in European discussions on evaluation, we will highlight how concern-based evaluation contributes to a better informed public dialogue, discuss whose interests it serves and the role it has to play in a pluralistic debate.

¹ National Institute of Forestry, Agriculture and Environment, Paris / Montpellier, France.
Involving special need populations in the evaluation of public services. Necessary input to adjust and implement public policies

Dr Renée Proulx
Postdoctoral Fellow
Department. of Social and Preventive Medicine
Université de Montréal, Canada

Co authors
Ruth Northway and Céline Mercier

This paper relates to the issue of involving special needs populations in evaluation processes that pertain to the development or implementation of public policies. Examples are taken from experiences involving people with intellectual disabilities (ID).

People with ID represent 1 to 3% of North American and of European populations. As several evaluation processes presuppose the ability to read and write and the mastery of a language that many people with ID might not possess, they are often excluded from evaluations. The symposium aims 1) to identify ways to facilitate and sustain the participation of individuals with intellectually disabilities (ID) in the evaluation processes and 2) to demonstrate how service users’ input can be useful to adjust and implement public services. Issues related to the recruitment, informed consent, procedures and interview formats are enlightened. Illustrations of how current approaches and methodologies can be adapted to allow people with ID to play active role are presented.

The identification and implementation of strategies to support the participation of individuals with an ID in assessment processes may result in the improvement of recruitment and interview strategies for other persons who have special communication needs, for example, persons who are illiterate and those with sensory limitations. This research theme falls within the wider perspective of political and organizational will for involving service users in policy development.
Abstract 125

Getting the views of parents and family carers of people with intellectual disabilities to inform service planning and development

Dr Ruth Northway
Professor, Unit for Development in Intellectual Disabilities
School of Care Sciences
University of Glamorgan, UK
Email: rnorthwa@glam.ac.uk

At present there are a number of changes which are impacting on the way in which services for people with intellectual disabilities are planned, developed and implemented. Developments such as person centred planning and direct payments have the potential to transform current patterns of service delivery. One key group of people whose views are important in relation to these developments are parents and family carers of people with intellectual disabilities.

The aim of this paper is to describe an evaluation process involving parents and carers of people with intellectual disabilities and to demonstrate how it is used to inform service planning and development. This project has been commissioned by a Parent Forum Alliance within one local authority area in Wales, UK since it was felt that parents should have the opportunity to express these views and that the information gathered should be used to inform service development.

Results from focus groups and interviews are presented. The data gathered has been analysed to identify emerging themes in relation to eight key areas: knowledge of the range of services currently available, satisfaction with such services, knowledge of planned service changes, hopes and fears in relation to such proposals, views concerning developments which seem necessary but are not being planned, and views concerning current mechanisms for involving parents and family carers in service planning and development. As the Parent Forum Alliance and the local authority are members of the project advisory group, findings are used to inform service planning and development within the local area.
Closer Co-operation Between Evaluation Societies - South - North and West - East

This will be a panel discussion with contributions from IOCE, IDEAS, EES and UKES who will each introduce the topic with a short description of their current activities and priorities.

We will work toward a set of practical recommendations.

The themes which we will address are:

- How can we as societies work together more effectively to achieve high standards of evaluation?
- What are the key themes for future activities? For example, working more closely with the donor community, voluntary sector and private enterprise
- How can we encourage better co-ordination between us and with other evaluation societies?
- Can we identify and agree any areas on which to co-operate in the near future?

Everyone involved in the future of international evaluation is welcome and encouraged to attend.
Many countries, one evaluation: critical lessons from the review of a regional project

Dr Daniel K. B. Inkoom
Department of Planning
Kwame Nkrumah University of Science and Technology, Ghana
E-mail: dinkoom2002@yahoo.co.uk

Co author
Dr Roland Msiska
Regional Project on HIV/AIDS
Johannesburg, South Africa
Email Roland.Msiska@undp.org

This paper highlights the lessons learnt from a Mid-Term Outcome Evaluation of the joint UNDP/Sida Regional Project RAF/99/022-Building Capacities for Reducing HIV Spread and Consequences on Development carried out in 2003. Conducted in 11 Sub-Saharan African (one IGAD, three ECOWAS, and seven SADC) countries by a team of four independent evaluators from different countries, the evaluation presents critical lessons for international evaluation and evaluation in developing countries.

The Regional Project aims at increasing the understanding and action on the interrelationship between HIV and development policies and strategies in sub-Saharan Africa, in collaboration with regional and national partners. This is with the view of achieving the following:

a) Reduce the impact of development efforts on the spread of HIV at national and regional level
b) Reduce the impact of AIDS on development efforts at national and regional level

This paper addresses three key questions, namely:

- What methodological and process issues were critical for the evaluation of a project in such a varied context?
- What impacts did the evaluation have on the structure, management and delivery mechanisms of the project? and
- What were the lessons learnt by all involved and how are these lessons being translated into the conduct of future evaluations?

The key issues the paper raises relate to the procedures, expectations and outputs of evaluation methodologies, and management of the evaluation process and how these were reflected in the countries visited.

The paper particularly teases out the interest generated by the evaluation in the evaluators themselves, the commissioning agencies, the UNDP country offices and local organisations dealing with HIV/AIDS and development.

The paper concludes that for such evaluations, two key issues are paramount: paying particular attention to context, and effective collaboration of all stakeholders.
Management and evaluation of regional development initiatives

Dr Peter Parker
Associate Professor of Business Anthropology
University of Kristianstad, Sweden
Email: peter.parker@e.hkr.se

According to the Swedish Agency for Economic and Regional Growth, NUTEK, approximately 10 billion Swedish crowns (more than 1 billion Euros) have been channeled annually into regional development projects in Sweden. EC structural funds, national and regional governmental agencies and private enterprises each contribute roughly a third of the total investment (NUTEK 2004).

Despite extensive efforts it has been very difficult to document results of this investment. An extensive independent review of the impact of structural funds in Sweden 1995-2000 finds no quantifiable affect whatsoever in relation to goal variables such as per capita income, employment rate, proportion of population of working age. Moreover, the report finds no support linking EC funding to any intermediary variables such as business dynamism, development of skills and labor market or the attractiveness of the region as a place of work or recreation (ITPS 2004). The lack of demonstrable results must surely be detrimental for the legitimacy of these regional development initiatives. It is argued in this paper that difficulties in achieving and communicating stem in regional development initiatives stem from a mismatch of means and ends.

The large scale bureaucratic procedures imposed in EC funding is singularly inappropriate for stimulating the kind of innovative development the programs seek to promote. We argue that existing theory on the management of complex projects provides great deal of insight into how regional development issues might achieve better results and be able to communicate them. Drawing on insights from the management of complex projects we outline an approach to the management and evaluation of regional development initiatives that address key problems of these initiatives. Ongoing work with regional development projects in Scania with these starting points is discussed.
Altogether now? Critical realism and the evaluation of health and social care partnerships

Helen Dickinson
Research Fellow
Health Services Management Centre
University of Birmingham
Email: hed496@bham.ac.uk

Partnership working has become of prime importance in the delivery of health and social care in the UK over the past ten years. Despite a growing recognition of the importance of inter-agency collaboration in the delivery of public services, most studies focus on process (how well we work together), not on outcomes for service users (Dowling et al., 2004). As such, aside from a notion that partnerships are necessarily a ‘good thing’, there is little evidence to show that such an important policy priority has had any real effect on service users. To redress this balance, this project involves the design of an online evaluation tool to measure the impact of health and social care partnerships. The design of this tool is derived from much of the literature on good practice in partnership working, and utilises a critical realist framework through which Theories of Change and Realistic Evaluation have been interpreted.

The intention of this tool is to surface the underlying assumptions about a partnership and how it is intended to affect user outcomes. These assumptions are then put to users to assess whether the organisation is delivering the “correct” outcomes and how successfully. From this we hope to identify some generative mechanisms of partnerships which may also be applicable in other contexts. Against this background, this paper discusses the role which critical realism may play in evaluating service delivery, the way in which this influenced the design of the assessment tool, and results to date.

Assessment of pharmaceutical needs within an East Midlands Health Authority

Dr Peter Rivers
Head of Research & Evaluation Services
Faculty of Education Health and Sciences
University of Derby, UK
Email: p.rivers@derby.ac.uk

Judy Staton
Research Associate
University of Derby, Uk
Email: Judy.staton@tiscali.co.uk

This study was undertaken to support planning for implementation of the new pharmacy contract, specifically those parts relating to 'enhanced' and 'advanced' services, within an East Midlands Health Authority.

The objectives were to:

1. establish an accurate and comprehensive picture of the range of non-contracted services currently provided by pharmacies within the catchment area of five PCTs
2. identify any additional services community pharmacists consider they might offer after the introduction of the new contract and their justification for such services
3. place these findings in the context of PCT and national priorities

Method
A structured questionnaire was designed for telephone interview to collect categorical information describing 'enhanced' and 'advanced' services offered by a single pharmacy. Pharmacists and PCT staff were asked to prioritise the enhanced services they thought the PCT should fund and to justify their views. The need for pharmaceutical services was considered for electoral wards with the highest prevalence of morbidity for certain critical conditions (e.g. 'circulatory diseases').

Results
A response was obtained from all 107 pharmacies via branch managers or head offices. A comprehensive list of all enhanced and advanced non contracted services was presented in tables and bar charts showing the number and percentage of pharmacies a) offering a given service / method of funding, b) plans to provide a given service with or without PCT funding.

Conclusions
Pharmacists' views on the services that merited funding were fairly consistent across the district. The highest priority was assigned to: blood pressure testing, cholesterol and heart disease clinics, followed by diabetes screening and monitoring, weight management advice and smoking cessation. For each PCT there was a substantial 'gap' between the services that pharmacists were providing and what they were willing to offer if additional PCT funding was made available.

References
Problems of policy evaluation in a European multi-level governance context: the case of active labor policy in Belgium

Bart De Peuter
Catholic University of Leuven, Belgium

Co author
Joris De Smedt

Within the European multi-layered governance system each of the increasingly intertwined governmental levels is subject to structural change and/or policy-related and financial modernization trends. However, these simultaneous reforms do not end the existing horizontal variance within each sub-European level concerning structures, competences and resources. Structural complexity, horizontal diversity within the lower levels and scattered competences over the different tiers of government are the main characteristics. Seen from a European perspective horizontal differences between countries remain, and within most of the EU member countries differentiation also exists between states, between regions and between local governments. This differentiated reform context entails in particular challenges in terms of capacity and coordination across governmental boundaries with regard to vertical policy evaluation. This paper addresses these issues in the field of rural development, within the federal country Belgium. The EU policy of rural development supports agriculture as a provider of public goods in its environmental and rural functions, and in their development.

With this paper we address the problem of policy evaluation in a European multi-governance context. In Belgium the Flemish state level is the pivotal tier of government in terms of responsibilities, but it operates within a federal context and under a European framework of standards. The Flemish government also has to rely on the cooperation of local governments in order to implement its policy priorities.

This case presents a typical snapshot of the structural complexity and intertwined competences between different governmental levels which entail difficulties for policy evaluation. From this case the paper will deduce generic preconditions which have to be fulfilled and critical success factors with regard to vertical policy evaluation across levels of government. From this perspective the paper contributes to the transfer of knowledge and problem-solving with regard to intergovernmental policy evaluation issues in similar circumstances.
Evaluations for the EU strategy for better regulation

Dr Per Mickwitz
Senior researcher
Finnish Environment Institute, Finland
Email: per.mickwitz@ymparisto.fi

The paper will discuss some insights gained through the practice of evaluation in relation to the European Union strategy for better regulation. The EU strategy for better regulation is largely on the political agenda as a means to improve competitiveness and the conditions for more growth and jobs, that is, it is part of the efforts to implement the Lisbon strategy. However, it has also been justified with the need, "to win back the confidence of our [EUs] citizens".

The paper will discuss how evaluations of EU policies could contribute to the discussion about the basis for selecting which regulations to "to scrap, modify or codify", in the words of the better regulation strategy, and how to modify them. At the same time, it should be recognised that many important aspects of EU policies can only be revealed and analysed in a particular local context, considering how a EU regulation has been implemented and how it has affected actual practices and been affected by already existing practices. National and regional public administrations cannot be viewed as merely implementers of EU policies. At each level new policies are made, with aims that either complement or are in conflict with those of EU policies. The real challenge is, thus to develop evaluation so as to genuinely serve the development of multi-level governance. The paper will combine conceptual discussions with examples based on evaluations of EU environmental policy directives considering their interaction with national and local policies.
Program theory evaluation of public-private agreements on environmental programmes and projects

Gustav Jakob Petersson
Mälardalen Evaluation Academy, Sweden
Email: Gustav.Petersson@mdh.se  gustav_gandhi@hotmail.com

This paper deals with how to secure accountability in situations where a special kind of agreements is made between public and non-public actors. Such agreements with private business on policy formulation and on the implementation of such policies have been used in several European countries in order to increase energy efficiency and reduce CO$_2$-emissions.

More specifically, the problem dealt with here is that the evaluator needs a supplement to the program theory method in order to be able to make an evaluation that brings forth the information needed to help securing accountability. In this paper some starting points for such a democratic accountability evaluation are discussed.

However, the difficulties are many and the Swedish ECO-Energy Programme is used to exemplify that the theoretical problems discussed also exist in practices.
Process monitoring of impacts - towards a new approach to monitor the implementation of structural fund programmes

Richard Hummelbrunner
Consultant
ÖAR Regionalberatung, Austria
Email: hummelbrunner@oear.at

Process Monitoring of Impacts is an innovative approach for monitoring the intervention of projects or programmes. The method aims at identifying processes which are essential for the achievement of results and impacts and to collect data or information which are required in order to observe these processes. It builds on the basic assumption that inputs as well as outputs have to be used by someone in order to produce desired effects. Thus focus is placed on the actual use of inputs or outputs by partners, project owners, target groups etc., which is considered decisive for the achievement of effects and can be influenced by the operators of a project / programme.

Process Monitoring of Impacts has been tested in 2005 in the framework of on-going evaluations of Structural Fund Programmes in Austria, in particular projects and measures which have "soft" characteristics and whose results are difficult to assess with conventional, quantified indicators (e.g. support for R&D activities, business innovation, networks). And it is currently applied with the preparation and ex-ante evaluations of several Austrian programmes for the new Objective 2 "Regional Competitiveness" and 3 "Territorial Co-operation".

At the beginning, the paper outlines the need and rationale for a new monitoring approach, based on experience with present monitoring systems and the ideas put forth by the EU Commission for the new programming period. The need for monitoring processes - along or in lieu of indicators - is explained in some detail and Process Monitoring of Impacts is described as an alternative approach to current monitoring practice. Experience gained with pilot applications in Austria are briefly outlined and the utility of the method is compared to current monitoring practice. The final section specifies how Process Monitoring of Impacts can be applied with Structural Fund Programmes in the new programming period.
Making connections in Europe: the role of the UK National Audit Office in assuring effective use of European Union funds

Peter Gray
Director
National Audit Office
Email: peter.gray@nao.gsi.gov.uk

The proper and effective use of its €100 billion budget is one of the most significant and contentious issues facing the European Union (EU) today. Evaluation and audit regimes are in place to monitor and report on how the money is used both at national and EU level. This paper looks at the role of the National Audit Office (NAO), the UK’s state audit institution (SAI), in respect of the use of EU funds. It examines its responsibilities and the ways in which the NAO acts to secure appropriate use of funds within the UK and, more generally, across an enlarged Union.

The paper examines the NAO’s activities in supporting the development of sound financial management in the EU. It will show how it does this on its own but also through its formal and informal connections with European institutions such as the European Court of Auditors and the European Commission, as well as with SAIs in other countries. Activities include direct audit work of the structural funds, collaborative studies with other SAIs, and involvement in multilateral forums such as EUROSAI. It also covers how the NAO is helping the development of financial management and performance audit capacity within newer EU countries and aspirant members of the Union through twinning and other arrangements. Finally, the paper will cover how the NAO has sought to explain the workings of the European budgetary system to a non-specialist audience through regular public reporting and by communicating key EU financial management matters to the UK Parliament in its reports on the main findings of the audits of the European Court of Auditors.

Overall, the paper will show how the sharing of expertise and knowledge is essential to the development of the European project and how the NAO’s connections are helping strengthen the use of EU funds.
Delegation, cooperation, and evaluation: how to solve problems of coordination in structural fund programs. Findings from Comparative Case Studies in the South of Italy

Dr Mita Marra
Researcher
National Research Council, Italy
Email: Mita.Marra@issm.cnr.it

European cohesion policies are increasingly relying on grassroots networks tapping into tacit knowledge and participatory decision-making processes. Regional governments delegate their decision making power to local institutions and require policy evaluation with the assumption that local agents possess both contextual knowledge and political legitimacy to integrate different policy measures in a cooperative fashion. Delegation of decision making power as well as systematic evaluation are presumed to minimize the unintended or conflicting outcomes emerging, for instance, when environmental protection and infrastructure building are not designed consistently to local contextual needs nor are these pursued through a cooperative effort of local networks of actors. Different agents, including resource users and government agencies try to work together to resolve shared dilemmas of coordination, as an increasingly common alternative to centralized institutions. Coordination consists of managing interdependencies among multiple individuals or organizations involved in the overall program or project management.

Several studies classify different types of coordination mechanisms, including standards, evaluation arrangements, hierarchy, targets or plans, slack resources, vertical information systems, direct contact, liaison roles, task forces, and integrating roles. Other ways of classifying coordination include formal impersonal, formal interpersonal, and informal interpersonal; non-coordination, standards, schedules and plans, mutual adjustment, and teams; task-task, task-resource, and resource-resource coordination; vertical and horizontal coordination; coordination by programming and by feedback; and coordination by standards, plans, and mutual adjustment. Building upon a current field research in four regions of the South of Italy, this paper examines how coordination occurs across local development programs, which are embedded within multilevel governance structures and relations.

The paper presents a number of cases of local collaborations in which large numbers of local actors representing a wide range of contending groups have worked out agreements for integrating structural fund programs and evaluated them. In some circumstances, specific coordination mechanisms encouraged consensus building offering all relevant groups the knowledge and skills needed to participate in evaluation processes. In other circumstances, though, delegation of decision making power opened the door for opportunistic participation, making evaluation just a bureaucratic compliance act.
**Outcome Evaluations - a promising form of result-oriented evaluations of development interventions?**

**Dr Thomas Winderl**  
Monitoring & Evaluation Consultant  
Email: thomas@winderl.net

The United Nations Development Program has piloted Outcome Evaluations for the last few years now. This seems a good time to assess if this new approach to evaluate development interventions is a useful form of result-oriented evaluations.

Outcome Evaluations are evaluation that cover a set of related projects, programmes and strategies intended to bring about a certain outcome. They assess how and why outcomes are or are not being achieved in a given country context and the role UNDP has played. The starting point is the outcome rather than a specific set of programs or projects (like in a cluster evaluations). The strategy is to work backwards from outcome to assess a number of variables.

Outcome evaluations assess four standard variables to assess: a) whether an outcome has been achieved or progress made towards it, b) how, why and under what circumstances the outcome has changed, c) UNDPs contributions (outputs) to the progress towards or achievement of outcome (incl. soft assistance), and d.) UNDPs partnership strategy in pursuing the outcome.

This type of evaluation also tries to capture soft assistance normally not included in project evaluations. Soft assistance consists of policy advice and dialogue, advocacy and brokerage/coordination services. Soft assistance is an often overlook area of development intervention but important input for the achievement of results.
Evaluating networks as a tool to drive policy change

Richard Longhurst
Senior Research Fellow
Institute of Commonwealth Studies
University of London, UK
Email: longhurst@clara.co.uk

Peter Wichmand
Head of Design
Evaluation and Documentation
International Programme for the Elimination of Child Labour (IPEC)
International Labour Office Geneva, Switzerland
Email: wichmand@ilo.org

In the international development world, the formal network has become the modern organisational form. Networks have become a very popular means of generating and disseminating knowledge and encouraging changes in behaviour. International organisations want to use them not only as a means of exchanging information and as a tool for advocacy but also to carry out research, build capacity and induce policy change. Networking in itself is a key activity; most people find informal networks very important for their work. What networks can achieve has often been exaggerated but when correctly conceived and implemented they have many benefits, although their maintenance costs and timeframe are often underestimated.

A review of the literature shows that common factors in the success of external networks are: a common purpose with clear objectives and talks and focus from shared need for action, non-hierarchical structure, clear benefits for all members, critical mass of members, adequate resources and long term commitment, and visible sponsorship.

This paper, authored jointly by evaluation commissioner and evaluator, reviews the evaluation of a project which intended to stimulate national, regional and global networks, creating them where necessary, supporting them where they existed already, of technically competent and committed institutions which could act as an ongoing force for child labour action. There were three networks, in the areas of development policies, hazards child labour policies and gender, all constituted in different ways.

The paper discuss the results of the evaluation, and provides some general ideas on how to evaluate networks, including the evaluation of projects that create networks (as this project did) versus evaluating networks as such, showing how networking, knowledge and capacity building interact, how networks at different levels are linked, the challenges of building evaluation into networks from the beginning.
Internal and external needs for evaluative studies in a multilateral agency: matching supply with demand in UNEP

Dr Segbedzi Norgbey
Chief of Evaluation and Oversight
United Nations Environment Programme, Kenya
Email: segbedzi.norgbey@unep.org

Co author
M. J. Spilsbury

We have entered an era where development interventions financed via the public purse are subject to intense scrutiny by funding agencies. Competition for funding is increasing and the expectations of funding agencies have shifted from the production of outputs towards an insistence that their contributions generate measurable positive impacts in the real world. Demands for accountability and transparency are commonplace and this is motivating an increasing number of institutions to invest in a rigorous portfolio of evaluations, - a trend that is set to continue.

However, the relative demands for, and the utility of, different types of ‘accountability-oriented’ and ‘learning-oriented’ evaluative products remains unclear. Intuitively, one would expect that higher standards of accountability and demonstrated benefits from prior investment (e.g. from ex post impact assessment) should increase the likelihood of maintaining or increasing funding levels. Nevertheless, informal discussions with funding agencies suggest that the linkages between evaluation findings and any subsequent resource allocation patterns are often weak.

Thus, many organisations supported through donor financing face the dual problem of satisfying donor demands for accountability and securing continued funding. A first step in addressing these problems is to understand the types of evaluative evidence that satisfy accountability demands and those that increase the likelihood of future funding.

This study explores how evaluation findings inform and influence donor funding decisions, and how evaluation findings are used within the organisation. The study is based on a survey of UNEP Governing Council representatives, UNEP’s donor agencies and senior UNEP managers. The survey examines preferences for different types of evaluation approaches and methods, and their perceived credibility, reliability and utility in relation to the resources required to produce them. The findings would reveal the relative importance of different evaluation approaches and highlight important tradeoffs e.g. the need to satisfying donor demands for accountability whilst meeting management needs for operational improvement. The study will attempt to highlight the likely level of influence of evaluation findings on resource allocation decisions.

The immediate objective is to increase the understanding of UNEP’s key stakeholders in terms of their perceived needs for evaluative products. Such information will help ensure that the allocation of available resources to evaluation activities is optimal. The survey findings would also have wider relevance for organisations that are accountable to a range of donor agencies and be useful for evaluation networks / ‘communities of practice’ associated with development assistance agencies.
Evaluation of Spanish cooperation aid policy: the need of a reticular approach

Carmen Vélez
Evaluators
Instituto de Desarrollo Regional, Espana
Email: carmenvelez@idr.es; anacirera@idr.es

Ana Cirera
Instituto de Desarrollo Regional, Espana

The evaluation process of the Spanish Cooperation Policy started in the middle of the 90’s. The Law launched in 1998 implied the reform of the intervention instruments (director’s plans, annual plans, sector strategies and countries strategies), the improvement of the project management mechanism, the design of a systematic evaluation methodology and the empowerment of the coordination and decisions bodies of the State.

Nevertheless, this system has not achieved a consolidate strategy in terms of efficacy, efficiency and quality of the aid and resources distribution to eradicate poverty and underdevelopment. Far away, it’s a more violent, intensive and complex planet problem.

The cooperation aid programs works like a net where some of the actors have a privilege status of power and some of the others have not the legitimacy to participate in a horizontal scheme. This is one of the principal characteristic of this policy, to have an asymmetrical relationship between developed countries, which financed the programs, and the receptor countries.

For this reason, it became necessary to look for new instruments and rules to share competences in the decision making process among public administration, institutions and other actors. In addition, it is also a key issue to adapt the evaluation framework in order to include a reticular net approach that allows to all the stakeholders to be considered when evaluate effects and impacts of the public intervention.

Network evaluation system makes possible the identification of the actor positions and connection in the net, and verified the institutional development and capacity to solve its interest conflict and ways of interrelationship. With this perspective we can ensure that all point of the net engaged in the design, implementation and evaluation process are highly involved to contribute to the Development Objectives of the Millennium, modifying governance model of the international cooperation aid programs.
Committee evaluation in the OECD

Kevin Williams
Evaluation Coordinator – Council Secretariat
OECD, France

In the context of the OECD’s recent reform programme, the decision was taken to introduce a mechanism to evaluate the work of committees financed from its main budget, as part of an overall package that included the establishment of a biennial results-focused programme of work and a monitoring mechanism for reporting on output delivery, quality and initial or anticipated impact.

The approach taken by the OECD has been to establish an internal evaluation function, working under the auspices of a group of five ambassadors (the Evaluation Sub-group) nominated from the Organisation’s highest body, the Council. The evaluation function is situated within the Council Secretariat and as such is not organisationally dependent on the administrative and substantive services which have an interest in the outcome of evaluations.

The stated aims of what the OECD calls In-depth Evaluation are:

- Providing lessons for the future
- Reinforcing transparency and accountability
- Informing Council decisions on resource allocation and on mandate renewals and appropriateness

Integrated into the evaluation mechanism is a procedure for monitoring the take up of evaluation recommendations by concerned committees, again implemented under the auspices of the Evaluation Sub-group of Ambassadors.

The Evaluation Function was established in September of 2005 and after elaborating of a broad terms of reference and methodology embarked on delivering a programme of six committee evaluations, decided by Council, before the end of 2006.

The first three evaluations were finished by March 2006 and an initial review of progress made and lessons learned was presented to Council in July 2006. Following on from the initial review, deliberations on the future direction of what the OECD are currently taking place.
The evaluation of European Community’s development aid programmes: How are thematic and country evaluations carried out?

Ivory Yong-Protzel
Senior Evaluator
European Investment Bank, Luxembourg
Email: yong@eib.org

Luc Lefebvre
Societe d’Etudes et d’Evaluations, Luxembourg
Email: luc.lefebvre@sciences-po.org

European Community aid to developing countries is quite significant and diversified. The institutional frame, the political and social context of the last 15 years as well as the changes in the management of the European Commission have all contributed to a rapid development of the evaluation of development aid and to the more explicit formulation of policy priorities. These policies are designed at sector level (transport, energy, private sector development, etc. for all developing countries) or at country level (defining a cooperation strategy with each country for all sectors) and are the object of synthesis evaluations – which presents a number of complex methodological challenges.

The paper aims at exploring the European Commission’s approach to evaluate Community aid to developing countries and its evolution towards more comparative and participative methods. The following issues are discussed:

1. A brief description of Community aid to developing countries - In terms of objectives and amounts in five different regions of the world
2. The evaluation of development aid as a recent and necessary evolution - Three factors have converged to make this possible:
   - The institutional frame of development aid
   - The window of opportunity, after the cold war, to analyse aid effectiveness
   - The trend within the European Commission towards a result-based management
3. A description of the European Commission’s synthesis evaluation of development aid:
   - Who is responsible and what is evaluated
   - How the approach is adapted to a complex situation:
     - The evaluation criteria used reflect European policy priorities
     - Interventions are evaluated against explicit policy statements
     - Effort is better focused using evaluation questions
     - Who carries out to these synthesis evaluations? Three partners - Independent evaluation with the active participation of services involved.
   - A real effort to disseminate and follow-up results with concrete instruments
4. Challenges to the future – and some tracks to face them
   - Accountability - To whom are development aid actors accountable?
   - The evolution of the nature of European development aid
   - The limited quality of data and the lack of norms
   - The high cost of these evaluations given the geographical dispersion of the interventions
   - The heterogeneity of the scope of such thematic evaluations.
   - How to face these challenges at a reasonable cost?
The denial of Politics in PRSP's Monitoring and Evaluation: experiences from Rwanda and Ethiopia

Dr Nathalie Holvoet
Post-doc Researcher
Institute of Development Policy and Management, Belgium
Email: Heidy.Rombouts@ua.ac.be

Dr Heidy Rombouts
Post-doc Researcher
Institute of Development Policy and Management, Belgium
Email: Nathalie.Holvoet@ua.ac.be

The new development paradigm strongly reconfirms and redefines the importance of M&E because of some newly incorporated principles. First, increased value is adhered to results orientation, iterative learning and evidence-based policy making. Realisation of these basic principles is conditional upon a strong and well-functioning M&E system. A second major principle of the new paradigm is an increased role and responsibility for the national governments in managing the entire M&E system. Donors are expected to dismantle their own 'parallel' M&E apparatus, to increasingly rely on and align to national M&E arrangements and systems. It seems that, when playing its renewed role, the international donor community largely focuses on technical issues of national M&E systems (quality of statistical systems, quality of indicators and targets, etc.), vigorously avoiding politically sensitive topics. However, some studies suggest that politics are part and parcel of M&E. Evaluations take place in a socio-political and economic context in which different stakeholders have differing - at times even competing - interests. This overall context determines the power relations between (and within) the parties and institutions involved. Both power relations and interests influence key evaluation issues. Since the new development paradigm incorporates new M&E challenges, the politics of M&E deserve to be studied from this new angle.

Power relations and interests at both the national and international level are expected to influence the overall M&E policy, its scope and organisational set up. Furthermore, it tends to influence donor's own assessment of partner countries' M&E capacity and related capacity building efforts. On the basis of an assessment framework and field material from Ethiopia and Rwanda, this paper furthers the understanding of the impact of politics on M&E within the context of PRSPs and unveils flaws in the theory of the M&E-PRSP framework. The paper identifies crucial elements and signals in the politics of M&E. It argues that persistent exclusion and denial of the importance and presence of political issues in M&E eventually risks undermining both M&E's functions of 'accountability' and 'feedback & learning'.
PRSP monitoring systems and the role of donors advantages and disadvantages of different forms of involvement

Dr Monika Zabel
M&E Specialist
HLSP, UK
Email: monika.zabel@hlsp.org

Henry Lucas
University of Sussex, UK

This paper aims to consider the reported activities and experiences of donors in PRS monitoring. It will seek to assess the current position and to provide some guidance as to possible ways forward. It was completed following a review of findings on eleven recent country studies and one regional study on PRS monitoring. However, the extent to which these studies focus on the role of donors in the monitoring process varies widely. The paper has therefore drawn extensively on the most interesting studies and added material from a range of other sources where considered helpful to the discussion. The note is structured as follows:

1. Introduction
2. Donor engagement with PRS monitoring
3. Budget support
4. Donor demands on PRS monitoring
5. Projects targeted to PRS monitoring
6. Support to data production
7. Support to analysis and dissemination
8. Discussion and recommendations

The next section provides an overview of the rationale for donor engagement with the PRS monitoring process and the forms which this might take. This is followed by examination of the special case of monitoring in the context of budget support. Whatever the aid modalities, and however cordial the relationship between donors and partner countries, the latter will tend to prioritise what they regard as donor demands for monitoring information. Section 4 therefore considers what demands it is reasonable for donors to make on the PRS monitoring system and section 5 the extent to which they should seek to influence the means by which those demands are met and to provide the necessary resources. The next section focuses on some of the specific approaches to data production that donors are supporting. Section 7 then addresses perhaps the least advanced area of PRS monitoring thus far, the intention to use the findings to engage the population, civil society and national researchers in policy debate. The final section provides some thoughts on current best practice.

Potential positive and negative characteristics of the various aspects of donor involvement in PRS monitoring are discussed, 'best practice' for each form of involvement is identified.
How evaluation guidelines, principles, codes of ethics and evaluation standards shape the connection between evaluation and society

Dr Thomas Widmer
Department of Political Science
University of Zurich, Switzerland
Email: thow@pwi.unizh.ch

Several national evaluation societies in Europe as well as numerous organisations have evaluation guidelines, principles, codes of ethics, evaluation standards or similar documents in place.

This panel - organised by the EES Working Group on Standards, Guidelines and Principles - will debate the question if and how these documents (can) contribute to the connection between evaluation and society. Based on their respective experiences with standards, guidelines, codes and principles, the contributors will discuss how these documents shape the connections between evaluation and society.

The panellists are people working in the field of guidelines, codes and standards on behalf of selected societies and organisations. Namely the following people will contribute to the debate:

Jean-Claude Barbier (French Evaluation Society SFE)
Wolfgang Beywl (Society for Evaluation DeGEval)
María Bustelo (Spanish Evaluation Society SES)
Sara Hector (European Commission, DG Budget)
Hans Lundgren (OECD, DAC)
Helen Simons (UK Evaluation Society UKES)
Thomas Widmer (Swiss Evaluation Society SEVAL).
Making multiple connections: developments in the value for money audit of the National Audit Office

Jeremy Lonsdale
Director
National Audit Office
Email: jeremy.lonsdale@nao.gsi.gov.uk

The National Audit Office is one of the largest producers of reports - more than 60 a year - evaluating the performance of public sector institutions, programmes and systems in the UK. Its value for money (VFM) work can be traced back to the nineteenth century but has developed considerably in scale and complexity since it acquired a statutory basis in 1983.

This paper traces the development of value for money work, highlighting how it has gone from a discipline influenced primarily by its audit roots to one which is now closely connected with a range of other disciplines - including evaluation, performance measurement, inspection and management consultancy. It highlights the growing debate about what constitutes 'quality' in VFM work, given the multiple influences and shows how the NAO has sought to balance the expectations of many stakeholders, including Parliament and Government. It also shows how external drivers - the development of evaluation practices, changing public expectations of public provision, and the desire of government to refashion the delivery of public services - have all shaped the VFM discipline to make it a diverse and perpetually changing activity.

The paper will link into a number of the conference themes, in particular, showing how the NAO has sought to use its work to contribute to the public debate about how public services can be best delivered in an environment favouring low direct taxation. It will also, briefly, contrast developments in the UK with wider European trends.
Evaluation of large initiatives of scientific research

Mary Kane
President
Concept Systems, Inc, USA
Email: mkane@conceptsystems.com

Professor William Trochim
Director of Evaluation for Extension and Outreach
Cornell University, USA

Government agencies and international organizations expend billions annually on scientific research initiatives to advance their fields of research and affect critical societal outcomes. Funders, policy and decision-makers and the public expect accountability for where, how effectively and how efficiently funds are expended; and evidence of progress on the problems research is funded to address. Multi-project research initiatives often require that researchers work interdisciplinarily or transdisciplinarily, organized through complex centers, collaborations, consortia or networks. These structures may be international or inter-agency, including private and public entities.

Such scientific research usually consists of multiple research projects undertaken to address a common critical issue, like the human genome project, international networks of HIV/AIDS prevention and treatment research, and initiatives on tobacco control and cancer, to name a few. In the European Union, such endeavors are often considered Science, Technology and Innovation (STI) evaluation. There is growing recognition of the need for evaluative feedback mechanisms to better understand and manage the structures that support such initiatives, and for overarching models and approaches that help deal with the complex systems issues involved in evaluating them.

This paper describes efforts to develop systems for evaluating research programmes in the U.S. National Institute of Allergies and Infectious Disease, and the U.S. National Cancer Institute. We describe a general approach to evaluation of large research initiatives guided by several principles. It must: be participatory and collaborative; include a comprehensive conceptual model of the initiative; use integrative mixed methods; link with existing information systems; encompass the life-cycle of the initiative; and be both standardized and adaptable. Methods are reviewed, including concept mapping, logic modeling, survey research, content analysis and peer evaluation of progress reports, bibliometric analysis, and financial analysis. We discuss preliminary results of several evaluations, look at lessons learned, and consider the major challenges in such work.
Evaluation - a tool for Diagnosis and Empowerment in the municipal educational system

Dr Anat Raviv
University Lecturer
Haifa University
Expert Adviser for Evaluation
Ministry of Education, Israel
Email: anatraviv@yahoo.com

Israeli society is composed of residents who for various reasons immigrated to Israel over the years from all over the world. Thus, the Israeli educational system has always, and especially in the last decade, had to cope with students of significant social, cultural and socioeconomic variance.

School is the main meeting point between children from different countries of origin; hence its important role as a major cultural bridging agent (Ben-Rafael, 1985; Peres, 1977; Wilcox, 1982). International studies found several key principles related to the successful integration of these students into the education system in particular and into society in general:

- Language Development – this approach maintains that the low achievements of students stem mainly from their severance from their mother tongue upon their entrance into school. As a result, these children lack the necessary linguistic skills to develop their thinking. It is therefore the duty of the education system and of the school to foster the language among these students at an early age.

- Ethnic Additive – this approach suggests intensive treatment of this population by adding study hours to develop core and linguistic skills.

- Cultural Pluralism, Cultural Differences – by including elements from the ethnic culture, and their incorporation in the various teaching and learning styles.

- Self-Concept Development – by focusing on the child's individual empowerment through his or her strengths.

One of the main factors found to have a very significant influence on the students' performance in school is intercultural differences in order to response these needs the education system in Israel has implemented various theories over the years, which did not prove to be successful (Peled, 1984; Peres, 1977; Shapira, 1977). In the last years one of the paradigms noted by Lynch and Banks (1986), "cultural pluralism" seems to be the most common system to deal with the problem.

These principles served as a basis for shaping an educational policy in one of the towns in the northern part of Israel, where 55% of the local population is composed of new immigrants from the Former Soviet Union and Ethiopia. These demographic features led the educational officials in this town to plan and lead a municipal action to evaluate the students in two basic subjects: language and mathematics, as a tool to cope with the cultural, social and socioeconomic variance of the learners' population. The goal of this evaluation was to identify the variance among the students and respond to their special needs.

The evaluation and assessment process is composed of external appraisal of the entire system twice a year and internal appraisal at different junctions throughout the year.
The municipal evaluation plan includes all the students in the town's elementary schools (N=2069) beside pupils in the first class of school.

The outcomes of systematic evaluation serve as the basis for decisions made at the different levels. The evaluation findings form the basis for developing intervention programs at the municipal, school, class, and individual levels. This municipal program creates a significant change in the perception of the individual needs and to the development of a unique individualistic approach and student empowerment, while breaking the vicious circle which characterizes low self-efficacy.

The study will demonstrate the evaluation process and the development of insights and intervention programs at the different levels, which have lead to a significant improvement in the students' achievements and in their integration into the education system, as a basis for their successful learning in the high level education in the future.

This model could be used also in other countries to respond to the needs of many immigrant students, who find it difficult to integrate into school and society at large.
An evaluation framework for assuring the quality of community language schools

Dr Tim Wyatt
Partner
Erebus International, Australia
Email: twyatt@tig.com.au

This paper describes an innovative approach to the development of an evaluative framework for assuring the quality of community language schools – an important sector emerging within the Australian educational landscape.

As the global population becomes more mobile, an emerging need for governments and non-government organizations is to ensure that migrants can contribute to the development of a cohesive community, rather than become marginalized or alienated from mainstream society. Arguably, respecting and nurturing cultural diversity is one of the keys to ensuring the connectedness of contemporary societies. An important strategy for promoting and sustaining a diverse cultural identity is through maintaining the linguistic heritage of all members of the society.

While languages other than English are taught in mainstream schools in Australia, many ethnic communities have established their own “schools” specifically to teach their home language and culture in what are seen as more authentic contexts. These ethnic schools usually operate after regular school hours or on weekends, using mainstream school premises that are made available at low or no cost to community groups. There has been a rapid expansion of the numbers of students enrolled in such schools over the past 20 years or so. Nationally, more than 100,000 students now take part in ethnic school language programs.

Despite the potentially significant contribution ethnic schools may make to both individual students and the broader society, until recently they have not operated within a consistent and coherent policy framework. The evaluation framework discussed in this paper is primarily intended for assuring the quality of community language schools. The process used to develop the framework respects the principle of strong local autonomy but provides information that can both drive local improvement and state and national accountability requirements, and because it draws on the expertise of all stakeholders involved in provision of community language learning, has resulted in an outcome that is widely accepted. This experience provides an example of how cooperative evaluation processes can build critical connections that contribute to the achievement of positive social and public policy outcomes.
A fidelity scale for assessing gender-specific substance abuse treatment: rationale, development and utilization

Professor Norman Layne
Professor of Sociology
Department of Sociology & Criminal Justice
UNC Pembroke, USA
Email: norman.layne@uncp.edu

Co authors
Misty Fulk
Director, CASCADE Services
Community Choices, Inc.
Sherry Edwards
Chair, Social Work Department
UNC Pembroke

Over the last 20 years, many substance abuse treatment models, modalities and interventions have emerged that focus on the needs of female substance abusers; collectively, these treatment techniques have become known as "gender-specific" treatment (GST) and are part of broader efforts to address a male bias in mental health and medical treatment. Many treatment programs advertise themselves as addressing the needs of women; however, there does not seem to be a consensus regarding treatment modalities and outcomes for GST programs. The type of treatment consumers receive from programs that call themselves gender-specific varies widely.

This paper identifies the common elements utilized by GST programs and synthesizes them into specific organizational and treatment elements. Each element is conceptualized and one or more 5-point scales are proposed that capture the essence of each element. The fidelity scale that has emerged from this process provides a systematic framework for assessing the degree to which a treatment program is providing GST. The data collection process triangulates information from five different sources.

The proposed fidelity scale has been tested and refined by gathering data from gender-specific treatment programs. The scale elements and the results to date of utilizing the scale to assess the delivery of GST are covered. Some of the benefits of developing a GST Fidelity Scale include the following: 1) Distilling out the key components of gender-specific treatment from a diffuse body of research and clinical reports; 2) Identifying the components of GST which are the most critical in producing successful outcomes; 3) Assessing a program's progress in implementing GST; 4) Providing feedback to administrative and clinical staff on what they are doing well and what they can improve; 5) Legitimating the resources needed to provide GST.
Factors affecting evaluation influence in the public sector: A case study of a government agency and its evaluations

Brad Astbury
Senior Researcher
Collaborative Institute for Research, Consulting and Learning in Evaluation at RMIT University
Australia
Email: brad.astbury@rmit.edu.au

There has been considerable discussion recently among evaluation theorists and practitioners about the merits of embedding the role of evaluation in public agencies. This emerging body of work represents another way of thinking about evaluation use that, arguably, more accurately reflects the various influences that evaluation research can have in different contexts. The ideas emanating from this growing literature, in particular Mark and Henry's (2004) schematic theory of evaluation influence, offer a way of organising future research on the impact and influence of evaluation.

This paper presents the results of research into evaluation influence in a government agency in Australia. The study explored factors facilitating or inhibiting evaluation and its influence and involved: (a) an integrative and theoretical review of the relationship between key concepts in the evaluation capacity building and evaluation influence literature; (b) an examination and synthesis of evaluation studies that were undertaken by the agency; and (c) interviews with senior managers to explore their perceptions and experiences. The paper concludes with a discussion of lessons learnt for enhancing the role of evaluative evidence in organisational decision-making processes.
The adaptation of the Centers for Disease Control and Prevention evaluation framework in public health in a New Zealand context

Dr Janet Clinton  
Senior Lecturer  
School of Population Health  
University of Auckland, New Zealand  
Email: j.clinton@auckland.ac.nz

The aim of this paper is to describe the use and adaptation of the Centers for Disease Control & Prevention (CDC&P) framework for public health. The CDC&P evaluation framework is accepted as the seminal document for use in large scale community based evaluations.

Three distinct evaluations were conducted in which the CDC&P framework was used as the foundation for the evaluation plan. The first evaluation was a large community wide programme to prevent type II diabetes; the second was a national programme to stimulate work force development and increase access and services to Pacific island communities in New Zealand; and the third was a workplace health promotion programme developed by a large non-government organization.

On each occasion the CDC&P framework required adaptation to local and the national NZ context. As a consequence a number of patterns emerged that illustrated that the CDC&P framework while providing a great foundation for evaluation, there was a need for adaptation on three levels. First, a specific cultural view needed to be included: the evaluation needed to be seen through a cultural lens. Second, a feedback loop was required at every stage of the evaluation framework. Third, a learning environment and development of evaluation capacity needed to be incorporated into the model.

This paper concludes that the CDC&P framework is valuable as a foundation for evaluation of health interventions, but it is imperative that the model be adapted to suit the needs of the local community and indigenous populations.
The Street Wardens Evaluation: lessons in applying experimental designs

Dr Kevin Marsh
Economics Professional Lead
Matrix Research and Consulting
Email: kevin.marsh@matrixrcl.co.uk

In 2003 Matrix RCL was commissioned by the ODPM to conduct a 4-year long ex-post evaluation of the Street Wardens Programme (SWP), a programme aimed at improving the liveability of local neighbourhoods.

This paper uses the experience of evaluating the SWP to consider the validity experimental approaches to evaluation, and how a closer interaction with the realist evaluation perspective can enhance the validity of the experimental approach.

The evaluation of the SWP included both a process and structure evaluation and a quasi-experimental impact analysis. The process and structure evaluation identified the factors that determine the success of the SWP and concluded that the SWP had generally been well implemented. However, the quasi-experimental impact analysis failed to isolate the impact of the SWP and was unable to conclude whether the SWP had improved residents’ perceptions of liveability, due to heterogeneity in the context of implementation of SWP, in the characteristics of those receiving the intervention and in the process of implementation of the SWP.

Conventional wisdom suggests that using a randomised controlled trial (RCT) is the best way to overcome potential bias. Drawing on the lessons from the SWP evaluation, this paper considers the ability of the RCT to control for the different types of heterogeneity and produce unbiased outcomes. It is argued that, in its own, the RCT only produces unbiased outcomes in the presence of heterogeneity of participant characteristics. However, this is not the case in the presence of heterogeneity of context or treatment. Two developments to experimental research design are proposed to ensure the internal validity of outcome measures:

1. consideration of context and treatment process heterogeneity requires that the RCT is supplemented with multiple-regression modelling, and

2. the determination of the appropriate treatment process and implementation context parameters within such models requires a closer integration of process and structure evaluation and impact evaluation.
Evidence-based policy making: fact or fiction?

PricewaterhouseCoopers Symposium

How did Superman defeat the Ku Klux Klan? What do schoolteachers and sumo wrestlers have in common and why do crack dealers still live with their mothers?

Even in today’s complex, multi-layered, globalised world, awash with confusion, distraction and even deceit, these and the other riddles of public life have real answers – but reaching them demands a new way of searching through the clutter.

Good data and sound evidence lie at the heart of effective public policy. Deciding if a gun is more dangerous than a swimming pool, may be a less pressing government priority than deciding if universal childcare today will reduce inner city crime tomorrow; but the evidence-based principles are the same.

PricewaterhouseCoopers LLP argues that finding the right perspective can unravel even the most complex phenomena; that sound evidence, proper evaluation and good analysis trends to deliver reliable and robust policy.

The PricewaterhouseCoopers’ symposium intends to explore, through a series of cases studies, how statistics, economics and interpretation have assisted policy making. Chaired by Dr David Armstrong, a partner in PwC’s Research, Strategy and Policy Group, three speakers will present recent PwC investigations, followed by a panel-led discussion on the issues.

The three case studies are:

The role of rigorous evaluation at project, programme and policy level

Reinout van Brakel
Senior Advisor Evaluation
PricewaterhouseCoopers LLP

In June 2002 the European Parliament and the Council introduced the Employment Incentive Measures (EIM) Programme. The programme is focused on policy makers, practitioners and researchers, and aimed to drive forward particular aspects of the European Employment Strategy.

In 2004, PwC was commissioned to undertake an independent evaluation of EIM; this has included a mid-term evaluation during the third year of the programme, an update of the mid term evaluation in 2005 and an ex-post evaluation during the programme’s last year (2006). This presentation discusses how a rigorous top-down and bottom up-evaluation leads to better policy making.
Real time evaluation - its impact on shaping public policy

**Brendan Miskelly**  
Senior Consultant  
PricewaterhouseCoopers LLP

Government introduced The City Academies programme in 2000 as a key element of New Labour’s school improvement strategy. The initiative focused on raising educational standards whilst maintaining a strong commitment to social equity and equality of opportunity. It aimed to turnaround failing schools where previous interventions had failed.

In February 2003, PwC was commissioned by the Department for Education and Skills (DfES) to conduct an independent, five-year evaluation of the Academies initiative. The aim was to assess the overall effectiveness of the initiative, in terms of its contribution to educational standards, and to examine the impact of key features of the Academies programme including sponsorship, governance, leadership and building design.

The evaluation regularly informs the DfES on the impact of the programme - positives and negatives. That gives DfES the opportunity to address any issues raised, change policy, with the aim of improving the programme and ultimately succeeding in raising educational standards.

Building peace without the evidence

**Jonathan Greer**  
Senior Consultant  
PricewaterhouseCoopers LLP

Measuring and evaluating the outcomes of post-conflict peace-building is a huge task; one that has challenged donor agencies as international development policy has shifted towards promoting and building peace in areas of conflict.

This policy shift follows an increased understanding that social and economic improvements don’t automatically eradicate community conflict. However, while sophisticated evaluation tools exist within the development field, they have not provided the solid evidence that captures intangible peace building and reconciliation outcomes. This problem has also been experienced in Northern Ireland/Ireland as funding bodies have struggled to agree on indicators that reflect the specific objectives of the PEACE II Programme.

Following a review of the academic and practice literature and consultations with stakeholders from donor agencies and the academic/evaluation sector, this presentation explores the difficulties in evaluating peace building outcomes and reviews international best practice methodological approaches. The presentation concludes by drawing together key principles of lessons from each of the methodologies and outlining key steps for conducting evaluations of peace building interventions.
Learning from multiple evaluations

Developing an evaluation strategy for health improvement policy in Scotland
Avril Blamey
NHS Health Scotland, UK

Developing a policy review process
Jill Muirie
NHS Health Scotland, UK

Building evaluation capacity and capability
Bron Macdonald
NHS Health Scotland, UK

Learning across multiple evaluations
Clare Beeston and Emma Halliday
NHS Health Scotland, UK

Chair and facilitator
Sanjeev Sridharan
University of Edinburgh, UK

In this symposium, we reflect on the issues raised by developing and implementing a more strategic approach to evaluation within an area of cross-cutting policy - population health improvement – in Scotland, a small country in Europe with a newly devolved government. Having been given the role of coordinating evaluation in the health improvement policy area, a key challenge has been to focus, prioritise and engage. The evaluation strategy has been developed with policy-makers within the Scottish Executive and is gradually involving a wider stakeholder group of those involved in policy implementation as well as the academic sector. The main aim is to improve the planning and delivery of health improvement policy by advancing knowledge and understanding about the effectiveness of policy implementation, by strengthening capacity and skills around programme planning, monitoring and evaluation and by developing a culture of learning, innovation and experimentation within which evidence, evaluation and review are integral parts of policy planning and delivery.

The symposium will include short presentations from Health Scotland’s Policy Evaluation team in order to encourage discussion about issues such as:

1. What are appropriate evaluation approaches for policy areas that comprise multiple programmes and policy measures?
2. How can we effectively link population health monitoring and target-setting to evaluations of policy implementation?
3. Finding processes to bring together learning across multiple evaluations
4. Developing a strategic approach to capacity building activities
5. Can we link the policy evaluation process to the identification of future policy priorities
6. How do we sustain the strategy?
Complementary and alternative methodologies in program evaluation

Professor Lee Sechrest
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: sechrest@email.arizona.edu

The designation "scientific" is not binary, i.e., either scientific or not scientific. Rather, scientific represents a continuum from "very strong" to "very weak." Good methodology-and good science-is the best methodology possible under the circumstances that surround our questions. Most of our evaluations of social programs will have to be carried out with methods less rigorous than randomized experiments, but that does not make those evaluations nonscientific.

Program evaluators need to think more carefully while framing their questions so that plausible rival hypotheses can be identified and, as far as possible, ruled out. Strong theory puts fewer demands on methodology and may obviate the need for randomized experiments as illustrated by several evaluations of interventions. Administrative or natural events may occur in such ways that they can be exploited by researchers to provide evidence for social policy, again as illustrated by actual evaluation studies. More effort needs to be put into "patching up" relatively weak methodologies or failed experiments, again as can be illustrated from our own work and that of others in diverse area.
Evaluation of Medical Marijuana Policies in the United States

Dr Dennis Gorman
Associate Professor
School of Rural Public Health, Texas A&M Health Science Center, USA
Email: gorman@srph.tamhsc.edu

J.Charles Huber
School of Rural Public Health, Texas A&M Health Science Center, USA

Empirical analysis of data pertaining to the effects of changes in marijuana laws is rare.1 While there has been some research on the effects of changes in marijuana laws in Europe and Australia, there are indeed very few such studies in the United States.2,3 Thus, there is a clear need for empirical research to inform the decision-making process in this area of public policy. In the case of medical marijuana laws, critics have argued that these send a message to the public that use of the drug is acceptable and may even have beneficial health affects. If this hypothesis is correct, then one would anticipate greater marijuana use and attendant problems to follow the introduction of medical marijuana laws. We test this hypothesis by examining the effects of the introduction of medical marijuana laws in California, Oregon and Colorado on arrestees using data from the US Department of Justice’s Arrestee Drug Abuse Monitoring Program. Specifically we compare longitudinal trends in positive drug tests among arrestees in five cities in these states (Long Beach, San Francisco, San Diego, Portland and Denver) with 16 comparison cities in states that have not introduced medical marijuana laws. In the analysis we use both regression and time-series models. The presentation will discuss the findings of our analysis of these trends along with a discussion of the challenges of evaluating the effects of changes in laws pertaining to marijuana use.


Counterfactuals in evaluation: The role of comparison groups in evaluation

Sharon Millar
Programme Evaluator
The Health Foundation
Email: sharon.millar@health.org.uk

As evaluators we are constantly plagued by the question of counterfactuals and we have to decide which one of two positions it suits us to adopt:

Do we accept that counterfactuals will always exist and can never be fully controlled?

OR

Do we continue to explore ways and means of reducing their existence?

Bearing in mind the pivotal role that evaluation is increasingly playing in shaping and informing public policy, after deliberation, we at The Health Foundation decided that it suited us to make a sincere if modest attempt to address the counterfactuals. We opted to include a comparison group in our processes and whilst we were aware of the limitations of this methodology, perhaps most significantly that the 'ideal design can never be achieved as the perfect equivalence of a comparison group can never be fully achieved' [Treasure Board of Canada] we believed that as an additional but complementary part of a comprehensive evaluation, there would be some benefit to be had.

Whilst it is still early days, this paper presents the case for undertaking a comparison group, outlines the methodological approach adopted and shares some challenges that were encountered along the way.
Feasibility of conducting an RCT to evaluate a single intervention in a custodial or community-based setting

Jacque Mallender
Managing Director
Matrix Research and Consultancy Ltd, UK

Jack Cattell
Senior Research Officer
Department of Transport, UK

Simon Morris
University of Cardiff, UK

Ben Styles
National Foundation for Educational Research, UK

The benefits of using experimental methods such as Randomised Control Trials (RCTs) to investigate the causal effects of an intervention are widely acknowledged. However, it is also recognised that experimentation, as it is conventionally applied in outcome evaluation, poses significant challenges, not just in ascertaining whether a specific outcome is attributable to one intervention alone as opposed to other possible causal factors, but also in relation to ethical and legal issues that stem from applying the methodology. Nevertheless, when it comes to answering the question “Does an intervention work?” it seems that there is no alternative method that can rival a well-designed and implemented RCT in terms of its robustness of evidence and in telling us what works.

Despite this, RCTs are relatively few and far between, particularly in measuring the outcome of public interventions in Europe. As pressure for increased accountability of public funds starts to mount, this may progressively change. This symposium brings together evidence and experiences of conducting RCTs in the UK from a group of evaluators representing the public and private sectors as well as academia. The aim of the session is to discuss what challenges successful RCT implementation and what might constitute a successful RCT implementation strategy. Difficulties and solutions are provided from previous studies. This includes a study of the feasibility of performing an RCT evaluation for an offending behaviour programme which the Home Office commissioned Matrix Research and Consultancy and the Jerry Lee Centre, University of Pennsylvania, to undertake between October 2005 and March 2006.
Using multi-method approaches to conduct evaluations in immigrant/migrant communities

Dr Zoe Clayson
President
Abundantia Consulting, USA
Email: zoeclay@abundantia.net

This presentation will compare two evaluations of complex, multi-site initiatives located in immigrant/migrant communities focused on improving community health in the United States of America. Three issues will be considered: selecting indicators for longitudinal study using both quantitative and qualitative data sources; constructing cross cultural terminology and research concepts; and bridging the divide between accountability and contemporaneous learning.

The selection of appropriate indicators to monitor within the time frame of the evaluations was challenging within these communities. Analyzing changes in large health-related data sets was further complicated by a desire to disaggregate data into geographic boundaries impacted by the interventions. In and out migration to and from particular neighborhoods within the initiatives presented another challenge. Information extracted from secondary data sources was augmented with primary qualitative data collection including conducting stakeholder interviews, and using photography and videography to describe change.

The understanding by immigrant communities of research terminology such as participation in the public sphere depended on the country of origin and cultural/political values of the constituency. For example, the interpretation of the concept “civic participation” varied widely from obeying traffic laws to involvement in informal networks from their country of origin. Considerable discussion was required with respondents to obtain intelligible information regarding levels of such participation. Surveys using scaling methodologies were less effective in garnering useful responses.

Finally, approaches to bridging the divide between accountability (fidelity to the model) and learning and changing during the course of the initiative (continuous quality improvement) will be presented. The use of a web-based approach provided scaffolding for achieving both objectives.

The three foci of this presentation indicator selection, cross cultural terminology, and frameworks for accountability and learning will contribute to the discourse regarding the design of evaluations in immigrant/migrant communities.
Social organization theory, families, and communities: Implications for evaluation science

Dr Jay Mancini
Professor of Human Development
Virginia Polytechnic Institute and State University, USA
Email: mancini@vt.edu

Social organization theory has been promoted as a significant way to understand the linkages between families and the communities in which they live. This theory has as its major components: Community Capacity; Social Capital; and, Network Structures. These components mediate between elements of the social structure and individual and family outcomes (one example being resilience). This theory has importance for understanding community contexts and their multiple influences on families.

This presentation seeks to elaborate social organization theory for the purpose of explicating implications for evaluation science, that is, program development and evaluation. The theory provides guidance for critical aspects of community and family life that are best aligned with prevention and intervention programs, and their accompanying evaluation strategies. The theory in particular raises questions of what aspects of communities might be evaluated, those that are related to specific programs, as well as those that are not program related. Recommendations for building and assessing community capacity are also included, since programs are limited in their larger community effects.
Street-level evaluation of social services

Dr Daniela Ferrazza
Researcher
RiSMes Department of Social Reasearch
University of Roma "La Sapienza", Italy
Email: danielafernazza@gmail.com

A recent Italian reform has advocated the creation of an integrated system in the delivery of social services. Evaluators are challenged to bear with the complexity of integration - either as integration between actors (i.e. public and third sector or state and regional levels) or as integration between services (i.e. social and health services) - on top of the intangible and contractual nature of the social services.

An evaluation conducted on the local service system in Rome focussed on the service of social secretariat, that epitomizes the principles of the reform.

Do the social workers really put "integration" into practice? In order to answer this question, we chose a methodological approach that can be summarized in the "street level evaluation" formula, which includes a direct observation technique, a participatory evaluation process and a peculiar attention for the practitioner's strategies, according to the assumption that integration is something people make.

There are different factors we were able to catch thanks to this approach and they regard basically those aspect of the social work that can’t be found in documents and declarations. First of all, organisational aspects about the informal division of labour influence very much the quality of the service. Then, the strategies social workers put in action to bear with the clients' requests highlighted the existence of at least two different kinds of outlook, named "pre-reform" and "post-reform" outlook. All this helped us understand how social workers from different services could cooperate, and overcome unpredicted obstacles.

Last but not least, during this process social workers became more and more confident in disclosing information that led to new suggestions for further evaluations.
Evaluation and the Elucidation of Professional Competence

Professor Roger Ellis  
Social and Health Evaluation Unit  
University of Chester, UK

Dr Elaine Hogard  
Social and Health Evaluation Unit  
University of Chester, UK

This paper considers the role of evaluation in exploring the nature of professional competence in social and health care and education. A three-pronged model of evaluation is outlined which focuses on outcomes, process and multiple stakeholder perspective. (Ellis and Hogard, 2006). Essentially these prongs address the questions: did it work? what happened? and what did they think? Within these three broad questions distinctive research questions and methods of gathering data to answer them are described for each of the three components.

In particular this paper focuses on methods to investigate the process of a programme. This part of an evaluation can describe, model and analyse exactly what happens in the delivery of a programme. It is argued that in social programmes and initiatives a key element is the process of interaction between professional and client. However this is still a relatively under-researched area with unanswered questions regarding the nature and effectiveness of interpersonal skills in, for example, nursing, social work and education. It is suggested that commensurately process evaluation is often neglected in programme evaluation and this neglect is exemplified through a consideration of key texts in evaluation. Through this relative neglect an opportunity is missed to describe and analyse the processes of professional client interaction and to relate these processes to outcomes.

Methods to describe and analyse interaction are outlined including observation, analysis of records and consultation with participants. Consultative methods are reviewed including critical incident analysis, Delphi technique and constitutive ethnography. A distinctive consultative method-reconstitutive ethnography—is described and exemplified in an evaluation study undertaken by the authors in education and health care. In this study (Ellis and Hogard, 2003) the processes of clinical facilitation in Nurse education were explored using a unique blend of critical incident analysis, constitutive ethnography and Delphi technique. This culminated in a description and modelling of clinical facilitation, which has influenced practice and policy.
The role of the evaluator and social change

Dr Janet Clinton  
Senior Lecturer in Evaluation  
School of Population Health  
University of Auckland, New Zealand  
Email: j.clinton@auckland.ac.nz

Co authors  
Dr Paul Brown  
Faith Mahony  
Dr Rob McNeill

The aim of this paper is to describe the process undertaken to develop the evaluation of the “Let’s Beat Diabetes” (LBD) programme. A major issue is how this journey has challenged the role of the evaluator, traditional methodology, and the interaction between evaluation and research.

LBD is a five year, district-wide plan for Counties Manukau, New Zealand aimed at long-term sustainable changes to prevent and/or delay the onset of Type II Diabetes, slow disease progression, and increase quality of life for people with diabetes. It incorporates ten distinct but interrelated areas of activity. A ‘whole society, whole life course, whole family’ approach is integral to the plan.

The evaluation framework and implementation plan was designed collaboratively with a community evaluation reference group and uses the CDC&P (1999) evaluation framework as its foundation. The aim of the evaluation is to conduct an independent programme evaluation, to guide and promote organizational development of LBD, to foster a learning environment, and to build capacity in evaluation and research workforce. The evaluation team were seen as true partners in the bringing about societal change.

There were a range of stakeholders and they had many evaluation requirements. These needs ranged from understanding the individual changes in health, to organizational development, and finally measuring population change in diabetes. To address these many needs required challenging traditional research and evaluation methods and creating new approaches. There were also important consequences for the communication to ensure feedback to all stakeholders. This on-going journey for participants, theorists, practitioners, policy makers, and evaluators provides a number of challenges for evaluation and raised a number of issues that impacts on the world of the evaluation.
Ensuring and assessing the quality evaluation reports

Svend Jakobsen
European Commission
DG Budget

Co authors
Timo Takkula
Eduardo Zapico

The quality of an evaluation is one of the most crucial factors for its results to be used. This was confirmed by a recent study on the use of evaluation results within the European Commission.

The quality of the final evaluation report depends on both the commissioning body and the contractor who carries out the evaluation study (around 85% of the evaluations are outsourced). The assurance of appropriate evaluation quality involve factors both at the level of the organisation (e.g., capacity building measures, evaluation standards, procedures) and at the level of the individual evaluations (e.g., measures to ensure adequate qualifications of the evaluator, support and monitoring throughout the evaluation process). Control and audit can also supports evaluation quality by ensuring compliance with evaluation standards. The presentation will outline how this is organised within the services of the European Commission.

Particular emphasis will be on the Commission’s approach for assessing the quality of individual evaluation reports. Such assessments can help to: a) ensure transparency about the quality of the evaluation to all potential users; b) encourage evaluators to conform to the terms of reference and to professional standards; and c) enable organisational learning about the evaluation process.

Both auditors and the Commission services have recently reviewed the practices within the Commission concerning quality assessments. A working group of evaluation staff has revised and simplified the quality assessment criteria being applied and a set of indicators have been identified as a help to gauge each individual quality criterion. The resulting revised reporting format, revised assessment criteria and a list of indicators for each criterion will be presented.

This presentation will be of interest for managers and contractors concerning outsourced evaluations (including consultants planning to tender for evaluations in the Commission) and it is also relevant for self assessment of internal evaluations.
Evaluating EU legislation and beyond

Nicolas Dross
European Commission, Belgium

Matteo Sirtori
European Commission, Belgium

Alan Zoric
European Commission, Belgium

The growing interest in the evaluation of the effects generated by European legislation raises challenging issues for evaluation practitioners. Using examples, two aspects in particular will be presented and discussed during this symposium:

(i) how to investigate the effects generated by legislation

(ii) arrangements for involving stakeholders.
Methodological Challenges in Impact Evaluation: The Case of the Global Environment Facility

Jos Vaessen
Researcher
Institute of Development Policy and Management
University of Antwerp, Belgium

Co author
David Todd

In this paper, we explore some of the methodological challenges that evaluators face in assessing the impacts of complex intervention strategies. We illustrate these challenges, using the specific example of a pilot impact evaluation study of one of the six focal areas of the Global Environment Facility; its biodiversity program. The paper discusses how theory-based evaluation can provide a basis for meeting some of the challenges presented.
Evaluation of CGIAR

Elliot Stern
Evaluation Consultant, UK

No abstract available
Strengthening national evaluation capacity development: the strategic role of evaluation professional organizations

Jean Quesnel
Director
Evaluation Office
UNICEF, USA

Greater professional recognition is in the making. When one draws a map of existing and emerging evaluation associations, groups and networks, it becomes evident that evaluation is increasingly being valued. From the seminal purpose of sharing experience and mutual learning, has grown a more professional rallying, as evidenced by the growing consensus about evaluation deontology in the various evaluation associations.

The paper will present the evolution and progress of the Evaluation Professional Organization worldwide, and will focus on the strategic role they are having in developing national and local evaluation capacities.

Finally, the paper will show that the success or failure of Evaluation Capacity Development depends greatly on three conditions:

- The awareness and appreciation at the governmental decision-making levels of the importance and necessity of evaluation – in other words the existence of demand of evaluation.
- The institutionalization and meaningful integration of the various evaluation function in the government machinery at national, sectoral, programme/project, and sub-statal levels.
- The development of human and financial resources to support a professional, dedicated, and effective cadre of evaluators and evaluation managers.
- The professionalization of evaluators and the recognition of evaluation as a discipline of work.
Continuity and coherence on the subject of evaluation quality

Kim Forss
Andante - tools for thinking AB, Sweden

Symposium Presenters:
Stein-Erik Kruse
Stefan Molund
Agnes Mwaiselage
Anna Nilsdotter
Evert Vedung

Since the Joint Committee on Standards published “The Programme Evaluation Standards” (1994) there has been an increasing attention to quality in the evaluation community. Various approaches to quality assessment have been developed (Forss and Carlsson, 1997), many evaluation societies have published guidelines to quality, or even standards. Recently the OECD/DAC (2006) made new standards public.

The emphasis has mostly been on various definitions of what constitutes quality. The standards define attributes of what constitute desirable qualities of either evaluation processes or evaluation reports). In a recent publication Mayne and Schwatz (2005) discuss various approaches to quality assurance. However, there have not been that many studies that have actually applied standards to evaluation practice, thus;

- There is not much practical experience of how standards can be applied. It is quite a long step from the formulation of desirable attributes of evaluation to a systematic use of them to ascertain quality.
- Consequently there is no overall picture of how “good” evaluations tend to be in respect of, for example, the OECD/DAC criteria, and hence no sense of were major shortcomings are, what efforts to improve quality should focus on, or areas where perhaps evaluations are generally quite good.

During 2005 the Swedish International Development Cooperation Agency (Sida) initiated a review of evaluation quality. The present authors developed a model to assess the quality of final written evaluation reports. The quality criteria were derived from the Program Evaluation Standards, the OECD/DAC guidelines, as well as from other sources. Based on these criteria we formulated indicators and developed scales for assessment. A population of 96 evaluation reports was identified, and we made a random sample of 34 reports that were analysed in detail. A report on evaluation quality was delivered to Sida in April 2006; this was a draft and the final report will be delivered in May this year.
Stitches, patterns and finished quality garments: a realistic evaluation of the complexities of formal parenting support

Karen Whittaker
Senior Lecturer
Department of Nursing
University of Central Lancashire, UK
Email: kwhittaker1@uclan.ac.uk

co author
Professor Sarah Cowley
King’s College, London, UK

Concerns about early life experiences and later well-being and behaviour have been recognised internationally and acknowledged by global organisations such as the WHO and UNICEF. From this a number of UK health, education and social care policies and programmes have arisen affecting families and children. New programmes however, arrive with old challenges of how best to deliver services and optimise outcomes. In essence new modes of delivery do not relieve institutions of their responsibility to assure that quality can be demonstrated and specific standards can be met.

In this paper we draw on evidence from a realistic evaluation of formal parenting support and argue that to recognise quality there is a need to understand how things work. Moreover in socially complex programmes, understanding the process of service access and use is the preliminary step towards defining realistically achievable standards.

The exemplar is a qualitative inquiry into the experiences of those using and delivering services. The nature of experiences is explained using the Pawson & Tilley (1997) realistic evaluation premise that outcomes (O) develop when mechanisms (M) are fired in favourable contexts (C) and previously common experiences become changed. In this study Bandurian (1986) social learning theory was applied to identify how C, M and Os, like knitting stitches, combined to form patterns and ultimately finished garments. The patterns illustrated the interplay between ‘person’ and ‘institution’ and came together cumulatively to arrive at institutionally defined outcomes, such as: uptake of training for employment or even reduced risk of child neglect/harm. The developed theoretical perspectives about the nature of formal parenting supports, offer a clearer understanding of how support was used. This provides a truer appreciation of which service ingredients reflect quality and the different forms standards may take as markers of sound institutional provision.
Making connections with the community: the opportunities and challenges of evaluating community-based health promotion projects

Ross Conner
Director
Center for Community Health Research
University of California Irvine, USA
and
President of the Board of Trustees
International Organisation for Cooperation in Evaluation

The World Health Organisation has been instrumental in fostering community-based approaches to health, with ‘health’ broadly defined by the communities themselves to encompass many social and economic issues. This approach has taken root around the world, but the evaluation of these projects had lagged. This EES-UKES conference presentation will discuss the situation in the USA and the author’s experiences in collaborating with partners from diverse US communities, including Chinese, Korean and Latino, on planning, implementing and evaluating projects involving various focuses, including cancer control and HIV/AIDS prevention.

Drawing upon his experiences as well as those of others working in ‘healthy community’ projects, the presentation will include examples from several representative projects, as well as a discussion of opportunities and challenges that these projects have for evaluation. In particular, the presentation will focus on the important but difficult issue of measuring changes in both individual-level and community-level intermediate- and long-term outcomes. For example, these projects usually have individual-level, intermediate-term outcomes tied to the content of the program (e.g., improvements in cancer-related attitudes and behaviors) as well as changes related to the collaborative approach (e.g., feelings of personal empowerment). Individual-level changes, however, are just one step in the process: Changes at the community-level are critical if the intermediate and long-term outcomes and impacts of community health improvement are to be achieved. These include changes in community capacity building and, ultimately, in community-wide social determinants of health such as economic and education levels, which address the root causes of inequities and inequalities.

This presentation will address the conference’s general theme of “critical connections” and the importance of nurturing relationships in undertaking evaluation, which are particularly important in community-based health evaluation. This presentation relates to Stream 5, with its focus on connections between evaluation and the improvement of social and public services.
Facilitating Self evaluation

Dr. Duncan Holtom

Senior Researcher
People and Work Unit
32 Monmouth Road
Abergavenny
Monmouthshire
NP7 5HL

E-Mail: duncan.holtom@btconnect.com
Tel: 01873-850-975
Mob: 07970-187-962
Fax: 01873-855-190
Web: http://www.peopleandworkunit.org.uk

The Health Promotion Division with the Welsh Assembly Government is seeking to strengthen the self-evaluation of activities supported by Health Alliances and the Health Promotion Voluntary Sector Grants Scheme, through the work of ‘critical friends’. The approach involves a voluntary sector organisation, the People and Work Unit, with experience of both external and self-evaluation, providing on-going support to grant recipients to enable them to improve their self-evaluation.

In order to structure discussions with grant recipients, the People and Work Unit has encouraged grant holders to reflect upon the relevance of the three core roles of evaluation – ‘accountability’, ‘development’ and ‘knowledge’ (Chelimsky, 1997) to their work and the ways in which they might evaluate it.

A key challenge for the work has been exploring how to evaluate the impact of often relatively small-scale health promotion interventions. In particular, the perceived problems in assessing impact, such as the difficulty of establishing causal relationships between interventions and health outcomes, has been used as a justification to either not attempt to assess outcome at all, or to only focus upon measuring outputs, such as the tangible services that the intervention provides. Yet without ‘hard’ evidence of positive health outcomes, there is an understandable reluctance to invest scarce resources in health promotion interventions or work that addresses the wider determinants of health.

The paper will discuss the different ways in which support has been provided, grant recipients' responses, and how some of the challenges common to health impact assessment have been addressed, in order to critically assess the effectiveness of this type of support.

References

Evaluating programs that use policy approaches to reduce underage alcohol problems in the USA: recommendations that can guide implementation

Dr Linda M Bosma
President
Bosma Consulting LLC, USA
Email: linda@bosmaconsulting.com

To reduce underage alcohol and tobacco use and related problems, several public health efforts in the United States have attempted to change formal and informal community policies to change the environment in which young people make decisions about substance use. One such program, Communities Mobilizing for Change on Alcohol (CMCA), has been designated as a Model Program by the US federal government's Substance Abuse and Mental Health Services Administration (SAMHSA) and is being implemented by numerous groups throughout the USA. CMCA utilizes a grassroots community organizing approach to enact policy changes which reduce young people’s access to alcohol. Such policies include compliance checks, training employees of retail establishments where alcohol is sold, and monitoring areas where young people can obtain alcohol.

Policy and community organizing approaches represent a new paradigm for many public health organizations that more typically implement educational efforts through school-based programs or community education. Such programs have typically been evaluated with student surveys to assess age of onset of use, past 30-day use, and perceived risks from use of alcohol. These evaluation methods do not adequately assess environmental prevention approaches as they do not assess the effectiveness of the groups that are organized to pass these policies or the community perception of those efforts. It is essential that groups evaluate these programs with appropriate methods that assess the actual nature of these interventions.

This paper is an examination of the community organizing process of three groups implementing CMCA. Research methods included key informant interviews, participant observations, and document analysis to examine policy outcomes, individual outcomes of participants in the community organizing process, and community outcomes related to the impact of the CMCA groups. Based on this analysis, recommendations will be made for the evaluation of CMCA and similar community organizing programs that can inform implementation.
Frail older people’s experiences and use of health and social care services

Markus Themessl-Huber
School of Nursing and Midwifery
University of Dundee
Email: m.themesslhuber@dundee.ac.uk

Gill Hubbard
Department of Nursing and Midwifery
Cancer Care Research Centre
University of Stirling

Aims: To evaluate older people’s experiences and expectations of services and the consequences for service provision, service development and research.

Rationale: A large amount of resources have been invested in providing services for frail older people who experienced multiple hospital admissions. However, their own views are underreported.

Method: Semi-structured interviews with frail older people were conducted in four Scottish Health Board areas to explore the context of emergency admissions and the use of extramural services.

Outcomes: Frail older people are high users of services but claim that services are not responsive to their main concerns: Meeting individual needs; maximising independence; and helping to live fulfilled lives. Services not catering for these needs are often cancelled or left in abeyance.

Conclusion: The same people who are targeted by care services are reluctant to engage with them. Care providers need to finally adopt older people’s priorities to provide them with responsive patient-centred care.
Evaluation theory, methods, and consequences: past problems, current solutions, and future developments

Dr Gary Henry
Professor
School of Education
University of North Carolina-Chapel Hill, USA
Email: gthenry@gsu.edu

Dr Melvin Mark
Professor of Psychology
Penn State University, USA
m5m@psu.edu

The Future of Evaluation Theory: Reflections and Projections
Melvin Mark

This paper discusses possible futures for the state of evaluation theory, drawing in part on observations of the past and current state of evaluation theory. Among the issues to be addressed: (1) Many classic evaluation theories implicitly emphasize different leverage points for influencing democracy; (2) Today there is advocacy of multiple, sometimes overlapping, sometime conflicting approaches; (3) There has been some shift from more pragmatic to more value-based evaluation theories. Drawing on such observations, a case is made regarding a desirable future for evaluation theory.

Evaluation Methods: Past Problems and Recent Developments
Gary Henry

Problems associated with evaluation methods, including designs, measures, data collection, and data analysis have discouraged some evaluators from using rigorous methods and inspired others to develop solutions. In this paper, new developments in methods including mixed method designs, imputation of missing data, propensity score matching, combining quantitative and qualitative data in the same analysis, power analysis, group random assignment, and new measurement techniques will be discussed. Examples from actual evaluations will be given to demonstrate the impacts of new techniques. Specific problems which are predicted for future development will be emphasized.

Use and Influence: Evaluation's Future
Melvin Mark and Gary Henry

We argue that “influence” is generally a more appropriate aspiration than “use” for public policy processes within democracies. At the same time, given the multiple functions and actors within democracies, and given that governments in practice involve mixed models of democracy, we suggest that evaluation needs to be attuned to targeted forms of influence (direct use and changing beliefs) and unplanned opportunities for influence. In addition, the mechanistic linking of evaluation results and mandatory or automatic forms of use is likely to increase, which increases the need to avoid inappropriate use.
Evaluating a model of program sustainability in health promotion

Brad Astbury
Senior Researcher
RMIT University, Australia

Dr Gerald Elsworth
Principal Researcher
Collaborative Institute for Research, Consulting and Learning in Evaluation
RMIT University, Australia
Email: gerald.elsworth@rmit.edu.au

We argue that sustainability in health promotion rests on common values that entail the equitable distribution of resources to present and future consumers and the continued achievement of health promoting outcomes. A realist theory of program sustainability that acknowledges these values is presented and evaluated.

Sustainability is viewed as a multi-faceted Outcome involving the maintenance, preservation and/or continued accomplishment of one or more of: (a) Individual participation during the project; (b) project activities beyond the funding period; (c) the implementing organisation itself; (d) the health promoting capacity of the organisation and/or community; (e) the service model exemplified by the project; (f) the ideas, principles and values promoted by the project; (g) the desired longer-term health outcomes.

It is proposed that sustainability occurs within two important Contextual Frames: (a) A multi-level ‘organisational’ frame of a project, nested within an organisation, typically funded by one or more agencies, themselves operating within various government policies, regulations and guidelines; and (b) a ‘level of constraint’ frame specified by the location of the program somewhere within the space defined by the extent of centralised control over the program, typically conceived as a continuum between ‘top-down’ vs ‘bottom-up’ approaches.

The model also distinguishes: (a) A set of Enabling and Constraining Conditions created by this context and the specific program rationale, guidelines and resources provided, (b) the Mechanisms generated as a result of the interaction between the context and the enabling and constraining conditions, and (c) the resulting Outcomes of the project at the levels of the project consumers, the project itself, the implementing organisation and the institution and/or policy community.

Data gathered from intensive case study of two food insecurity demonstration projects and more rapid appraisals of other health promotion projects utilising a modification of the critical incident technique illustrate and evaluate aspects of the model.
Sustainability of complex community interventions: riding the funding waves in changing organisational and policy environments

David Rankin
Research Fellow
Research Unit in Health, Behaviour and Change
School of Clinical Sciences and Community Health
University of Edinburgh, UK
Email: a.d.rankin@ed.ac.uk

The topic of sustainability has become increasingly important to funders, implementers and those employed by short-term funded public health programmes and projects. However, there are limitations to the growing literature on sustainability, much of it originating in the United States, including: absence of a common terminology, drawing on multiple research paradigms, and restricted focusus (specific projects designed to target a particular group, e.g. females and breast cancer screening rates, elderly people and heart health interventions). While insights from this literature are insightful, changes in UK public policy have led to the development of many short-term funded, complex community initiatives (CCIs), such as Healthy Living Centres (HLCs), which incorporate a range of objectives, partners and health topics. The sustainability of CCIs has received considerably less attention.

In this presentation I will provide an overview of the existing literature on sustainability before outlining the complexity of CCIs through drawing on data from a process evaluation (case studies) of six Scottish HLCs. Working as part of wider efforts to improve public health, CCIs are intended to address the complexity of underlying problems within target areas, through engaging with statutory, community and voluntary agencies alongside local communities. Findings from the process evaluation are used to examine how stakeholders from HLCs consider and explore options for sustainability, with reference to the dynamic systems in which the organisations are located. In charting the progress of case study HLCs toward attaining sustainability, this account will examine changed working patterns, organisational change within partner agencies, emerging policies, and continuing efforts to address community needs.
Evaluation as an instrument for capacity building development: the contribution of structural funds to enhance gender equal opportunities policies

Dr Flavia Pesce  
IRS - Istituto per la Ricerca Sociale, Italy

Dr Daniela Oliva  
IRS - Istituto per la Ricerca Sociale, Italy

Outline of the proposed content and argument of the paper:
This paper will focus on the direct (and un-direct) contribution of evaluation tools, within Structural Funds Programme, in terms of governance and capacity building.

Analysing in specific a sectorial theme, the paper shall discuss crucial factors that have become relevant for the success of the policies and actions to enhance gender equal opportunities within Administrations and the territory.

The analysis presented shall use different evaluation practices and practical examples the authors have experienced in Italy (at a national and local level). It will also, staring from the analysis of strengths and weakness, give lessons for the next programming period.

How the proposal relates to the Conference title and themes:
The proposed paper relates to the general conference theme of “the role of evaluation in public policy”.

In the proposed paper, the evaluation is proposed as an instrument that has an incremental purpose able to produce an organisational learning for different actors involved that improves efficacy, efficiency and quality of the policy itself and it considered as a useful support for new approaches to the theme of gender equal opportunities.

The paper, considering the complexity of Structural Funds, is also related to the specific themes of the analysis of multiplicity of institutional partners and stakeholders and the combination of policy tools trying to give an answer to the question of how can evaluation improve capacity building.
Building an evaluation culture: the case of the French drug policy monitoring system

Ivana Obradovic
Head of Studies
Observatoire français des drogues et des toxicomanies, France
Email: ivobr@ofdt.fr

This paper focuses on the origins and development of an evaluation approach to drug policy in France. It raises some critical issues that are important to consider if a useful evaluation is to be planned and carried out. It also offers the opportunity to reflect on what the criteria of a “useful evaluation” may be and how they shape the whole evaluation process.

As in many European or developing countries, evaluation of drug policy proceeds from a rather recent strengthening in the planning of drug policy. For more than three decades of specific public action against drugs, little evaluation work concentrating on the achievements had been undertaken, despite a number of policy changes.

In 1998, the cross-governmental authority in charge of all drug issues endorsed a brand new style of policy-making, with an emphasis on evidence-based programmes, relying on transparency and accountability. A three-year drug strategy was launched, setting priorities and targets, and commissioning, for the first time ever, an independent Observatory for its evaluation. A large amount of monitoring work has been achieved from 1998 to 2002 on drug prevalence, uptake of treatment and drug supply, without any formal evaluation framework.

The new Executive decided to sustain the reviewing of drug targets, along with a number of changes in the priorities. A second attempt has been made in 2002 to build a framework which would help monitor the strategic outcome of the current five-year strategy focused on a central thrust which is to reduce cannabis use by young people.

This contribution examines how the targeting and planning process initiated in 1998 has evolved since then. It also explores how specific opportunities connect with evaluation’s role and purpose in drug policy and what obstacles evaluation still faces. The very importance of political support in the evaluation process will be pointed out, together with an analysis of the current ways to work with commissioners to develop relevant and achievable evaluation of drug policy in France. Eventually, this presentation will consider the limitations of the existing evaluation system and provide some insight into its future development.
Evaluating regional knowledge society programme "eTampere" in Finland

Marja-Liisa Niinikoski
Managing Director
Net Effect Ltd, Finland
Email: marja-liisa.niinikoski@neteffect.fi

The eTampere Knowledge Society Programme was launched in 2001 and it ended up last year. The aim of the five-year programme was to transform Tampere into the world's leading city in the research, development and application of knowledge society by strengthening the knowledge base, creating new business activity and introducing new public online services. The programme was implemented by a variety range of actors in the Tampere region: universities, research institutions, the city, companies etc.

The eTampere programme consisted of six sub-programmes and it concentrated on three themes: 1) developing network services that ease citizen's daily lives and bringing them within reach of everyone, 2) strengthening the foundation of expertise in research and education, 3) creating new innovative business. The overall budget of the programme was approximately 131 million euros and it was carried out about 400 projects within the programme.

The ex post evaluation of the eTampere programme published this year focused on the evaluation of economical and social impacts of the programme at the regional, national and international levels. On the other hand one of the key evaluation issues was related to the applied programme model if the used model could be utilized in the implementation of industrial and innovation policies in the Tampere region in the coming years.

The main findings of the evaluation emphasize a thematic and cross-sectoral approach in the programme implementation and strong network management and governance tools when there is a wide range of organizations carrying out plenty of separate projects. Furthermore the integration of key actors already at the planning phase creates a strong basis for further co-operation when implementing an actual programme. The vision of the eTampere programme was strong and tempting and it raised interest in all sectors as well as nationally and internationally but accurate tools to integrate all relevant actors, especially SMEs into the programme were too loose in the first years of the programme's implementation. This fault also affected the results the companies seemed to gain from the programme. However the ambitious vision and targets of the programme raised debate and discussion and created circumstances for awareness raising and positive attitudes towards information and knowledge society at all levels.
Evaluation of changes in service delivery and organisational culture following the integration of adult mental health and social care services in Conwy & Denbighshire, North Wales

Sian Phillips
KTP Research Associate
Institute of Medical and Social Care Research
University of Wales, Bangor, UK
Email: s.e.phillips@bangor.ac.uk

This evaluation project is a Knowledge Transfer Partnership between the Institute of Medical and Social Care Research (IMSCaR) at University of Wales, Bangor and Denbighshire Local Health Board on behalf of the Conwy & Denbighshire Adult Mental Health and Social Care Partnership.

Information about Knowledge Transfer Partnerships is available at www.ktponline.org.uk

Mental health services in Conwy & Denbighshire have previously been managed by five separate organisations. The new Adult Mental Health and Social Care Partnership brings these organisations together with the aim of adopting a comprehensive, integrated and seamless approach to the care of working age adults with mental health problems in order to enhance the quality of the service its clients receive.

Despite co-location of health and social care staff in Community Mental Health Teams joint working has often been limited. Staff have associated more with members of their own professions. This barrier between staff has been sustained by each profession having its own management structure.

Operational differences also exist between all the Community Mental Health Teams. Clients have access to different services depending on historical and outdated procedures, local resources, their GP or postcode.

The new Partnership will address these differences by introducing an integrated management procedure, standardizing services across the two counties and working towards pooled budgets and shared resources.

A multi method evaluation strategy will allow both change and understanding of that change to be pursued. The approach will allow solutions to practical problems to be drawn from scientific enterprise, and conducted in collaboration with those who experience the problem. The active involvement of all stakeholders in assessing the effects of integration is a salient feature of the approach proposed in this project.

The evaluation of changes in service delivery and organisational culture resulting from the integration will inform future developments locally and nationally.

This evaluation will contribute to the accumulating evidence in the Cochrane Library for the effectiveness of alternative approaches to managing change in health and social care policy. It is also likely to inform the Welsh Assembly Government's decision making process around managing policy change in mental health and social care.
Evaluation in the Health Sector: the performance measurement system in the Tuscany context as a government tool

Sara Barsanti
Management and Health Laboratory
Sant'Anna School of University Studies and Doctoral Research

Co author
Professor Sabina Nuti
Sant'Anna School of University Studies and Doctoral Research

Tuscany has a strong public health care system organized in sixteen health authorities among which four are teaching hospitals, integrated with the Universities of Florence, Pisa and Siena. The twelve local health authorities that are not teaching hospitals deliver their services throughout community hospitals, outpatient departments, prevention services, with an approach that makes the evolution from the prevention concept to the health promotion concept clear and defining new operative forms open to the contribution of institutions and local social actors.

In Italy, Tuscany Region has tested and now is adopting an integrated model for performance measurement to which the regional administration, the local health authorities, and other shareholders may refer either in terms of indicators and shared responsibilities.

The aim of the Tuscany performance measurement system is to give a general outline for the management of the Health Care Companies, useful both for evaluating performance and for enhancing and promoting the results of the healthcare system achieved. In a Health Care Public System that strives to stand out for the process of cooperation and non competition between the subjects dedicated to providing health care services, it is important to plan and develop a transparent and shared system, capable of monitoring not only the economic-financial results of the Health Authorities, but their capacity of pursuing the aims of the Regional Health Care Plan as well. In this respect, the way the Institutions are managed and the output they obtain from the service delivery process in terms of clinical quality and citizens’ satisfaction gain great relevance. Therefore, through the performance evaluation system and the utilization of an essential number of indicators classified in six dimensions and represented in diagram targets, managers, their organizations and the regional healthcare system in their whole can learn and develop.

This evaluation system, used continuously and systematically at a regional level, it’s now a public policy tool that helps the Regional government to evaluate its strategic action.

Performance evaluation in healthcare services links policy to management and can be the mean through which output results may be oriented to achieve outcomes.
An Evaluation Needs Assessment in Public Health Programmes

Nicole Aerny
Jefe de Sección de Salud Pública
Consejería de Sanidad y Consumo de la Comunidad de Madrid, Spain
Email: nicole.aerny@salud.madrid.org

Co authors:
M Morán
V Civantos
C García
M Gil

The Public Health Department of the Community of Madrid launched in 2005 an evaluation unit that began its activities with an assessment of evaluation needs, as identified by the professionals responsible for public health programmes. The concept of evaluation that guides the action of the unit is based on the belief that participation is a key value for developing meaningful activity.

We will present the method that we use to gather information, including the design of a questionnaire applied in interviews with professionals, and the following process of estimation of the priority grade to give to each detected need for evaluation, considering the evaluability, vulnerability and relevance of the action to be evaluated and of the issue tackled by the action.

This process began assessing each need in a work group; to validate this first procedure, an individual assessment was then done, all results were put together to be compared and analysed, and discrepancies were cleared up. This is how criteria were thought out and a consensus about them was reached in the evaluation team. Finally, all assessed needs were distributed in three groups, according to the priority given to them as a result of this process.

In addition, an estimate was done of the complexity of each evaluation need in order to take into account the evaluation capacity of the team.

The last phase of this evaluation needs assessment consists of applying a political criteria: the Head of the Department chooses the evaluations to be carried out in the next two years by the unit, on the basis of a report about the evaluation needs detected, and their priority and complexity grades.
Evaluation of a pilot work based learning module for rehabilitation nurses

Dr Tracey Williamson
Research Fellow
Salford Centre for Nursing, Midwifery & Collaborative Research
University of Salford, UK
Email: T.Williamson@salford.ac.uk

Overall aim of paper
To disseminate findings from an evaluation of work based learning as an approach to developing the rehabilitation workforce.

Intended learning outcomes
Delegates will: 1) identify the advantages and disadvantages of a work based learning modular approach to development of rehabilitation staff; 2) gain insight into the application of the Philips & Stone (2002) model for evaluating training programmes

Three references

AIM: The purpose of this presentation is to share the findings and learning from an evaluation of a pilot work based learning project in the healthcare rehabilitation setting.

INTRODUCTION: A pattern was identified highlighting the inadequate preparation and professional development of much of the nursing workforce employed within eight rehabilitation departments across an acute hospital Trust. When explored further, a sense of feeling ‘overlooked’ and a lack of confidence and opportunity for studying at a higher education level was uncovered. In response, a one-year, elongated work based learning (WBL) module at Level 2 was developed in partnership with a local University to meet this group’s specific needs. The pilot comprised twelve rehabilitation nurses, yet influenced by the evaluation findings, there is an intention to roll out the WBL module across other disciplines and in an interdisciplinary manner.

EVALUATION STUDY: A comprehensive evaluation design was developed drawing on the work of Phillips & Stone (2002). Methods included interviews and collection of secondary data such as documents relating to anticipated learning outcomes, assignment marks etc. Comparative findings from interviews with the learners prior to and at completion of the WBL module, as well as six months on will be shared. These will be contextualised against findings from the individual and focus group interviews with other stakeholders including academic staff and Trust managers. The extent that stakeholder expectations have been met will be discussed. Delegates will benefit from insight into what worked well and what would be done differently in future development programmes for rehabilitation staff.

CONCLUSIONS: There is a scant evidence base concerning WBL in health and social care. Even less is known about the application of WBL approaches in the rehabilitation setting. This presentation adds to what is known and suggests future strategies for developing the rehabilitation workforce.
Innovations, Uncertainty, Evaluation and Learning

Professor Osvaldo Feinstein
Email: ofeinstein@yahoo.com

Innovations implemented in a context of high uncertainty, under which predictions are unreliable, require timely and careful assessment of results, drawing lessons from experience in order to facilitate learning. Given that this is a potential key role of evaluation as a discipline, one would expect that evaluation would be used for this purpose.

This paper discusses some of the possible reasons for the lack of a critical connection between evaluation and other disciplines in the evaluation of innovations, and provides suggestions on how evaluation could become more relevant in this area.
The Institutionalization of Synthesized Evaluative Knowledge: Adoption, Methods and Use

Professor Hanne Foss Hansen
Department of Political Science
University of Copenhagen, Denmark
Email: hfh@ifs.ku.dk

Co author
Olaf Rieper
University of Copenhagen, Denmark

In recent years several new institutions aiming at producing knowledge through summarizing and systematizing already existing evaluative knowledge have been established. Some of these are global such as for example the Cochrane Collaboration within the health sector. Others are national such as for example the What Works Clearinghouse established in the US within the educational sector. Common to these institutions are that they aim at ensuring that policy making and policy delivery is evidence-based. As such they represent a new way of organizing metaevaluation and knowledge synthesizing activities.

The paper sets out to analyse how this development initiated primarily in the US and the UK is adopted into the Scandinavian countries.

Also the paper analyses the assumptions regarding methods and use of this kind of knowledge production and discusses these in relation to policy making and professional practice. In addition to being crossnational the analyses is crosssectoral including the sectors of health, welfare and education.
Testing Relational Evaluation

Laura Yliruka
Researcher
National Research and Development Centre (STAKES), Finland
Email: laura.yliruka@stakes.fi

Kaija Hanninen
Researcher
National Research and Development Centre (STAKES), Finland

The aim of our presentation is to test the principles and explore the ideas of relational evaluation (Vataja & al. 2005) by examining closely two cases. Relational evaluation studies how the different actors produce and achieve the change by their collective action in which they use different resources. Relational evaluation is based on relational ontology and actor-network theory (Latour 1987).

Cases are two social work teams and their development activities. The evaluation bases on the triangulation of data that is collected during the development process. It consists of immediate feedback from developmental activities by each participant and focus-group interview done by researcher as well as follow-up questionnaire for both work-communities.

Both social work teams have employed the same method to their developmental work: Kuvastin (Mirror) – reflective self-assessment and peer assessment method. The idea of the method is to create an everyday learning and evaluative structure to social work practices that facilitates development of both individual and collective expertise. Both groups have been piloting the method in a developmental and research process conducted by researcher.

The cases have been chosen on the grounds that the results of data analysis have been quite opposite. The first social work team has stayed positive towards the developmental work and used method throughout the process and the other one has been quite critical towards both the process and the method. Interestingly both social work teams have however decided to continue to usage the Kuvastin –method after the pilot process has ended. Our interest is to ask what do we need to know in order to understand the differences between the dynamics of socio-material networks in both cases and how does this increase our understanding of relational evaluation.

Evaluation capacity in the EU's new member states and the new 2007-13 programming period

Jack Malan
Partner
Centre for Strategy & Evaluation Services (CSES) LLP, UK
Email: jmalan@cses.co.uk

With a substantial increase in European funding planned for the EU's 10 new Member States during the 2007-13 period, being able to properly evaluate programmes is more important than ever. At the same time, European Commission guidelines envisage devolving greater responsibility for evaluation than previously to national and regional authorities.

For the past two years, CSES has conducted an annual survey to examine the situation with regard to evaluation capacity in the 10 new EU Member States. The survey focuses on obtaining the views of evaluators and public officials on key practical and methodological issues. Comparisons are made between the changing situation in EU10 countries and evaluation capacity benchmarks for the 15 'old' Member States.

This paper will present findings and conclusions from the latest annual survey which is due to be completed in September 2006. Comparisons will be made with the results of earlier surveys to help identify trends. We will also draw on our experience of carrying out ex ante evaluation work for the new 2007-13 EU programmes in EU10 countries to illustrate findings and conclusions.

The paper will conclude by considering the implications from the research with regard to priorities for the future evaluation of EU programmes in the EU10 countries.
Complex tasks and ambiguous roles in transnational networks for innovation and transfer: lessons from evaluation for policy and practice development in EU partnerships

Philip Potter
Transnational Evaluations, Germany
Email: philip@transnational.evaluations.de

For evaluation to be a useful guide to innovation and transfer in policy and practice, it needs to be open to adopt ambiguous roles and set itself complex tasks.

In many EU programs, great importance is attached to the organizational form of the partnership as a vehicle on the one hand for generating innovation and on the other hand for promoting transnational transfer. Within these partnerships, evaluative activities are frequently led by a partnership member, hence constituting at the same time an evaluation internal to the partnership and external to the participating institutions.

Innovation: This dual internal/external role is appropriate for the evaluation of innovation, in which the evaluand is necessarily unstable, evaluative criteria cannot generally be determined in anticipation but must be progressively revised and refined, the evaluative objective is not so much the assessment of present effectiveness but rather of promising transfer potential, and for which qualitative, communicative, participatory and iterative evaluative procedures are in many contexts appropriate.

Transfer: The evaluative role of the evaluator as partnership member moreover often intersects for the purposes of transfer with roles of project coordination, knowledge management, capacity building, organizational development or even motivation and leadership. While this is seen by some as an ambiguity compromising evaluative identity, the argument in this paper is that this creative complexity creates opportunities for evaluation to engage more directly and creatively with policy development and practice transfer.

The paper draws on the empirical base of the author’s previous involvement in EU partnerships in the Socrates, Leonardo, Equal, Interreg, 5th Framework and EU-India ECCP programs.
Evaluation as a democratic intervention: Approaches for diverse populations

Dr Wendy M. Doughty
Partner
Emerging Directions Consulting Ltd, Canada
Email: wmd@emergingdirections.com

Birgitta Larsson
Emerging Directions Consulting Ltd, Canada

Conducting evaluations with culturally diverse populations poses both challenges and opportunities. Often persons accessing the program that is being evaluated face multiple barriers and vulnerabilities. Their realities and cultural setting can be far removed from those experienced by mainstream society and the cultural accepted norms. To introduce and conduct evaluations that are both meaningful to the commissioner of the evaluation and ensure buy-in and commitment from the population of interest can be pose challenges. Many of these challenges result from the evaluator being part of the mainstream culture and not fully understanding or appreciating some of the underlying values and protocols for conduct. This paper will present insights gained in conducting evaluations of programs accessed by persons for whom English is their second language, recent immigrants, adults with low-literacy levels, persons of Aboriginal heritage, and persons affected by domestic/family violence.

The paper will focus on preparing to undertake the evaluation. It will address ways to: (1) build engagement and commitment with stakeholders; (2) negotiate the evaluation approach; and (3) determine the scope and reach of the evaluation. The paper will also identify how successful evaluation processes and approaches can lead to vulnerable populations’ needs and service requirements becoming better understood by service providers. Case studies will be presented to illustrate democratic approaches to evaluation that engage stakeholders, build internal capacity and help ground the research.

The role of the evaluator in participatory and pseudo-participatory evaluation approaches will be explored. Ways to encourage those requesting program evaluations to consider more democratic approaches to program evaluation will also be discussed. Insights regarding the importance of determining proprietary rights of program evaluation results will be described. Finally, the role of the evaluator in developing a communication strategy in this regard will be shared.
Short-term, collaborative and post-conflict setting: the Bermuda triangle of evaluation practice?

Dr Valéry Ridde
Postdoctoral Fellow
International Health Unit
Université de Montréal, Canada
Email: valery.ridde@umontreal.ca

Co author
Sylvie Goossens

The consultation and participation component of the annual meta-evaluation of humanitarian evaluation action done by ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action) is always poorly rated.

We believe that such a poor utilization of participatory approach in humanitarian evaluation field is due to constraints such as limited time, resources, expertise, etc. This is why evaluations in humanitarian and post-conflict settings are mostly done through a directive approach, where external evaluators are controlling most of the process while stakeholders are the most consulted but rarely participants. Yet, we believe that not all evaluators are comfortable with that sort of practice and that it is possible to involve stakeholders in the evaluation process on the field.

The objective of this presentation is to demonstrate that, even in a post-conflict setting such as Afghanistan in 2003, a short term consultancy (limited to 3 weeks for resource constraint), allowed to adopt a collaborative and utilization-focused evaluation approach.

An implementation evaluation was done at the end of 2003 for a project implemented in the Eastern part of Afghanistan by a French medical Non Governmental Organization. The evaluation processes, coached by the external consultant, followed these steps: build and train a power balanced evaluation team, process the evaluability assessment and select the topics of the program under evaluation, define the evaluation questions, create evaluation tools, collect and analyse data, organize a lessons learned workshop and draft an action plan for decision making. We will discuss strengths and weaknesses of the process and we will draw some lessons from this case study.
The evaluation of the European Commission's Humanitarian Aid Activities 1992 to 2005

Peter Cavandish
European Commission
DG Humanitarian Aid

From its creation in 1992 the Directorate-General for Humanitarian Aid (DG ECHO) has been one of the world’s main humanitarian donors. DG ECHO has an established place in the humanitarian community because: it acts within a defined mandate; it is considered effective; and as it operates without a national political alignment, it is considered to add value both at EU and at international levels through the independent, impartial and neutral way in which it finances humanitarian aid. DG ECHO's entire life as a Commission service has been independently evaluated; by itself this could be considered to constitute an example of good governance in the public sector.

Since 2000 the Commission and most particularly DG ECHO’s management have undertaken many steps to transform DG ECHO. This has been in large part as a result of: the second evaluation of DG ECHO in 1998/99 - undertaken in accordance with DG ECHO’s Legal Basis, article 20 of Council Regulation (EC) No 1257/96 of 20 June 1996 concerning humanitarian aid.

The follow-up given to this second evaluation was undertaken at the same time as: the implementation of the Commission reform process; and the changes and new challenges affecting humanitarian activities world wide.

Among the matters addressed by management as a result of this second evaluation have been: the return to DG ECHO’s core mandate in the activities that DG ECHO finances; the introduction of the logical framework tool; the strengthening of strategic programming dialogues with key humanitarian partners; the reinforcement of internal systems and controls; the revision of DG ECHO’s principal contractual tool, the Framework Partnership Agreement (FPA); the introduction of ‘primary emergency decisions’ for fast track financing; the introduction of entry and exit strategies; the introduction of Global Needs Assessment and Forgotten Crisis Assessment tools; the introduction of thematic funding; the introduction of more formal policy and/or thematic papers on inter-alia, Disaster Reduction, LRRD, Children, HIV/AIDS; the augmentation of staff numbers, both at headquarters and in the field; etc.

DG ECHO's work over the period of 2000-2005 has now been independently reviewed by external consultants in the recent third major evaluation; their opinions can be made available to support the presentation.

The presentation will provide overviews on:
- the work carried out by the independent evaluators, including, how they were assisted, the consultation of stakeholders, within and outside of DG ECHO; reliance placed on the work of other professionals;
- the factors contributing to the take up or implementation of evaluation findings, including changes in the international humanitarian environment and Commission reform process; and
- the on-going role of evaluations at DG ECHO.
Poverty reduction interventions in developing countries: Evaluating lessons from KwaZulu-Natal Province (South Africa) in the context of government service delivery

Rishila Moonilal  
Chief Director, Monitoring and Evaluation  
Directorate of Macro Policy, Strategic Planning, Monitoring and Evaluation  
Office of the Premier KwaZulu-Natal Province South Africa  
Email: moonilar@premier.kzntl.gov.za

Simon Peter Opolot  
Office of the Premier KwaZulu-Natal Province South Africa

Andries van Staden  
Office of the Premier KwaZulu-Natal Province South Africa

Persisting poverty in developing countries is evidenced by the multi-dimensional deprivation of millions of people that are ensnared in spatial poverty traps. Particular geographical areas which remain disadvantaged include rural areas, urban slums, conflict zones, etc. For example, in the province of KwaZulu-Natal in South Africa, 74% of the 5.08 million people out of the total population of 9.4 million who live under conditions of poverty reside in rural areas. It is now universally accepted that people trapped in poverty are the cohort most likely to be multi-dimensionally deprived.

Thus, the thrust of programs aimed at freeing people from the dehumanizing effects of poverty has turned the developing world into a beehive of activity with billions of dollars being spent each year. Governments as well as local and international development agencies are investing resources (finance, equipment, personnel, etc) with the objective to eradicate poverty through achieving a particular set of socio-economic objectives.

But for all this unprecedented effort, are the poverty eradication interventions producing the intended benefits? What is the overall impact on the poor? Could the programs be better designed to achieve the intended outcomes? Are scarce resources being spent effectively and efficiently? We agree that impact evaluation can provide answers to these probing questions. Evaluation is particularly critical in developing countries where resources are scarce and for this reason every investment should maximize the reduction of poverty.

The knowledge gained from impact evaluation studies can also provide critical input to the appropriate design of future programs and projects. So, we discuss issues such as the availability and quality of institutions, capacity, etc in the developing world (citing examples from the Sub Saharan Africa region) for analytical rigor which is required in impact evaluation. Whose agenda should be served when poverty eradication programs and projects are evaluated? Does, and if it does, how does (how should) impact evaluation inform public policy development and implementation?
Results-oriented budgeting in Egypt

Doha Abdelhamid
Associate Professor of Finance
American University in Cairo, Egypt
Email: doha_ma@aucegypt.edu; abdelhamiddoha@yahoo.com

Professor Medhat Hassanein
Professor of Finance
American University in Cairo, Egypt

The paper will aim at documenting the success story of Egypt in the area of performance based budgeting monitoring and evaluation and shed the light on lessons learnt and anticipated implementation challenges ahead.

All governments, either in the Eastern or Western hemispheres, are under increasing pressure to produce results that would lead to tangible outcomes to their citizens. There is general recognition concerning the importance of an outcome focus for effective public management. Yet implementing a results-oriented approach has proved deceptively difficult in many nations in both the developed and developing world.

The Arab world and MENA region countries are now being faced by mounting pressures for undertaking bolder-style programs of reform in public financial management, as much as in the quality delivery of public goods and services. The political reality in the region dictates that the modes of reform vary from the standpoint that they may take shape out of homegrown or extra-territorial/cross-border initiatives. Irrespective of the mode, the need for reform is multi-dimensional and irreversible. With the interests of the Arab countries' grassroots expressed in the Alexandria Declaration in March 2004 and beyond, the option of reform blockage on the public democratic governance domain is not envisioned, given the current state of economic, political and societal affairs.

Having said that, the present paper should be read as a unique experience in which countries at a similar stage of economic development may take note of success areas and work on avoiding potential pitfalls given cultural specificities. Although this paper is intended to testify to the Egyptian success story on results-oriented budgeting, the experiment still has to be completed by expanding to interconnect all thirty-three government ministries on common understandings of the national development framework at a later stage. This particular aspect challenges the process of experimentation into the sustenance of a living organism.

The paper is composed of five parts, the first of which is an introduction. The second part discusses the salient features encountered in the transformation from input-based to outcomes-based budgeting in selected countries of the east and west. The third part is a discourse of the country-specific economic juncture in Egypt that instigated the irrevocable necessity to adopt results-oriented budgeting. This is manifested by reviewing the features of the present Egyptian budgeting system together with improvements undertaken since the multilateral reports’ preparation as demonstrated in the relevant sub-sectional footnotes. Section three, also, summarizes the two reports introduced that handle the readiness and possibility of adopting the new approach to budgeting for performance. The fourth section highlights the remarkable success of the efforts undertaken within the Egyptian ministries that showed a genuine commitment to participate in fiscal reforms which touch upon myriad development fields of public service delivery. The multi-faceted assessments portray both the World Bank’s and the Government of Egypt’s perspectives on project progress and prospective challenges. And, finally, section five, closes with lessons learnt and vision on the road forward.
Monitoring and Evaluation Reform under Changing Aid Modalities: Seeking the Middle Ground

Dr Nathalie Holvoet
Lecturer
Institute of Development Policy and Management, Belgium
Email: nathalie.holvoet@ua.ac.be

This paper grew out of our bewilderment with the insouciance with which some in the donor community seems ready to abandon accounting for the use of aid. If one listens to the rhetoric surrounding the new approach to aid, one gets the impression that most of the crucial accounting tasks must be swiftly abandoned by donors and left to recipient governments.

This paper does not question the underlying rationale for shifting towards recipient-led priority setting and control over implementation of aid resources, but argues that donors cannot let themselves off the hook that easily with respect to the accountability part of the equation. We argue that in most low-income countries such trust in recipient systems may be dubbed as overalignment, and that it is neither necessary nor useful. Our argument is however not that old style donor-managed monitoring and evaluation is the only or the best solution. For we are equally puzzled by the stubbornness with which some other donors stick to their old M&E in ways that contradict the new insights in aid effectiveness and hamper the emergence of national M&E systems. Why are positions so polarized and why is hardly anyone arguing in favour of intermediate positions?

This is what this paper sets out to do: we argue against a radical and rapid implementation of the new rhetoric, but also against a continuation of present accountability practices. Donors have a large and lasting responsibility in accounting for the use of aid funds, both towards the tax payers in donor countries and towards the targeted beneficiaries in the at best pseudo-democratic and poorly governed recipient countries. They should find new ways to remain firmly involved in M&E, ways that at the same time allow embryonic national M&E systems in recipient countries to grow and flourish.
The management of evaluation

Panel Session

No abstract available
Some characteristics of pre-evaluation of regional programmes in Hungary

Gábor Kulcsár
Analyst - Planner
VÁTI Kht, Hungary
Email: gkulcsar@vati.hu

In May, 2004 Hungary joined to European Union. Calls for projects in the framework of National Development Plan (NDP) has been launched and started the tendering and implementation phase of Structural Funds financed programmes.

According to the general rules of Structural Funds, indicators must be used in the implementation process to know more about outputs and results. Managing Authority (MA) coordinated to the system of indicators with guidelines and contributed the creation of a common indicator system. Huge demand has been arisen from decision-makers and stakeholders in the Monitoring Committee for a new type of evaluation on expected results in the earlier stage of program implementation.

On the demand of MA, consultants of VÁTI (intermediary body), have prepared special reports (pre-evaluation) on expected outputs and results of the programme in project selectioning phase by assuming project indicator values. The principle of the methodology was comparison between planned and expected results and estimation for impacts to promote project selection to favour more effective project.

Interpretation of indicators and building projects’ impacts for employment and equal opportunity has been provoked some methodological difficulties. Pre-evaluation took into account not only long-term employment, but short-term ones too. Describing direct and indirect impacts for employment by the case of some special types of projects (e.g. university campus or management training) were problematic also. Methods have been developed to estimate results and additional analysis were produced for every day use for increasing efficiency of implementation.

Above mentioned and other thematic reports were made regularly till up to finishing of contracting and contribute to enhance quality of implementation process.
Challenges for evaluation capacity building of local leader NGOs in Macedonia

Marija Nasokovska
Monitoring & Evaluation Coordinator
Institute for Sustainable Communities
Republic of Macedonia
Email: marijan@isc.org.mk

Macedonian civil society has been continuously strengthening in the last decade through diverse activities and programs supported by international donors. The focus was mostly on organizational capacity building, which has been indispensable for the growth of the sector. Monitoring and evaluation (M&E) were included in the strengthening activities (mostly training courses). However, the emphasis on M&E was never beyond the donors' interest in successful implementation of the projects' funded activities.

The Civil Society Strengthening Project's (CSSP) intention is to fill the gap in M&E that Macedonian NGOs lack in order to achieve a seat at the table and become legitimate participants in social and political decision-making. Thirteen NGOs were selected to participate and the project and will implement projects in different sectors and different geographical regions. The organizations are considered to be Leader NGOs, but there are variations in their levels of organizational development. Capacity building activities are provided parallel to the NGOs project activities and they are individually tailored according to the NGOs needs.

CSSP's evaluation capacity building activities are directed toward using M&E data as a management tool to improve individual project results and to help NGOs to be more accountable to themselves and their stakeholders regarding progress toward achieving their mission. The program design is flexible so it can respond effectively to NGOs needs. For instance, based on in depth interviews conducted with the Leader NGOs at the beginning of the project, planned the M&E group training as envisioned was rethought and now each organization will receive individual technical assistance in developing performance monitoring and evaluation plans for their projects. The interviews results demonstrated that evaluation capacity would be more significantly improved through "learning by doing" than simply to give the NGOs a theoretical background on "how to do it".

At this stage of implementation, CSSP is faced with numerous challenges in dealing with NGOs needs for building their evaluation capacity, such as organizational culture, motivation to commit resources, focus of outputs instead of outcomes and established systems for collecting, analyzing and using data for management decision making. This paper attempts to present experiences in overcoming the challenges and promoting local M&E capacity through the example of the CSSP's thirteen Leader NGOs.
Sustainability Impact Assessment of federal policy making: institutionalization of a bottom-up assessment process

Tom Bauler
Researcher
Université Libre de Bruxelles, Belgium
Email: tbauler@ulb.ac.be

Co authors
Paul-Marie Boulanger
Institut pour un Développement Durable, Ottignies-LLN, Belgium
An Heyerick and Erik Paredis
Universiteit Gent, Center for Sustainable Development, Ghent, Belgium

Belgium is currently promoting the institutionalization of an assessment process at its federal level, which is in the near future meant to evaluate policy proposals in an ex ante, participative and integrated way, before these are submitted to political approval. Such evaluation mechanisms exist in various configurations and depths since recently also on the European level (i.e. Impact Assessment), in Switzerland or in the United Kingdom (i.e. Regulatory Impact Assessment). The philosophy of the Belgian SIA is to leave the assessment process mainly in the arms of inter-departmental ministerial working groups, hence creating an institutional network and bottom-up coordination mechanism promoting the integration of Sustainable Development into (sector) policies. Necessarily, this bottom-up policy integration is intended to be complementary to the more top-down mechanisms of the federal SD-strategy.

Currently, the institutionalization of SIA for the federal level remains only very vaguely defined, and many different options coexist to organize and coordinate the SIA-processes at federal level, between departments, as well as between administration and political actors, as well as with civil society, stakeholders and experts. Based on a 2-year research project concerned with the methodological and procedural exploration of the future federal SIA, the present paper intends to discuss the effectiveness and efficiency of a number of these plausible and desirable institutionalization scenarios of SIA. The discussion of the scenarios will be both theoretical and enriched with the results of an inter-institutional discussion process.
The Impact of institutional information on investment's decisions: the case of FDI in Sweden.

Dr Federica Calidoni-Lundberg
Analyst
Swedish Institute for Growth Policy Studies (ITPS), Sweden
Email: Federica.CalidoniLundberg@itps.se

Co author
Patrik Arousell
Swedish Institute for Growth Policy Studies (ITPS), Sweden

In the last decades there has been a great deal of works dealing with the role played by institutions, and in particular public policy, in the case of market failures. There are mainly three types of market failure which are discussed in economics: the first relates to market power, the second to externalities and the third to imperfect information. Government intervention could then be motivated, from an economic perspective, in order to correct any of these failures. This paper focuses on asymmetrical information and supports the idea that market failures due to imperfect information problems do not require normative or regulative policy instruments but can be solved merely with the provision of a good level of information, capable of reaching the target group.

As Vedung (1997) states "there are three basic instruments that governments have recourse to: the stick (regulations), the carrot (economic means) and the sermon (information)." The sermon is the core topic of this article.

The first part of the work points out which are, according to the mainstream literature and to our opinion, the fields of public intervention where transfer of knowledge, communication of reasoned argument and persuasion are instruments good enough to solve a market failure and no regulations or economic means are required to lead to changes in people or firm's behavior. The instrument of information is less demanding than other instruments for it does not involve government obligation or material resources and it is, therefore, important to shed light on all the possible public areas in which government can reach the desired results "saving" resources and evaluate what impact the information could have on the recipients and on the market.

The second part of the article concentrates on the case of foreign direct investment in Sweden and illustrates the role of information provided by Invest in Sweden Agency (ISA) -the government agency assisting and informing foreign investors about business opportunities in Sweden- and how to measure the impact of such information on companies' decisions on whether to invest in Sweden or not.
Developing a methodology to assess the return on investment from professional development programmes: case study of The Health Foundation's leadership development programme

Dr Sarah Edwards
Programme Evaluation Officer
The Health Foundation, UK
Email: sarah.edwards@health.org.uk

Assessing the value for money of and return on investment from professional development programmes presents a challenge for evaluators, researchers and human resource managers. The need to assess return on investment comes from a range of internal and external drivers including the need to demonstrate that resources have been well used (public accountability) and that this expenditure has led to the desired impacts and outcomes. Through the case study example of The Health Foundation's leadership development programme evaluation this paper will explore how we are developing our approach to assessing return on investment using both quantitative and qualitative approaches.

The evaluation of the leadership programme is considering a wide range of process and structure, impact and return on investment questions. This paper focuses on the return on investment element of the evaluation and discusses how we are addressing the following questions:

What do we mean by the terms 'value for money' and 'return on investment'?
How can we capture the 'soft' impacts that are a key part of our return on investment?
How can we demonstrate 'impact' and sustained impact in the long-term?

The paper will provide an opportunity to engage with these debates and draw on the experience of those attending as we continue to develop our approach.
The evaluation of the new approach new solutions (NANS) Québec strategy to promote success for secondary school students in disadvantaged areas: lessons learned from the process evaluation

Dr Christian Dagenais
Professor
Montreal University, Canada
Email: christian.dagenais@umontreal.ca

The New Approach New Solutions (NANS) Strategy is the latest plan of the Québec government to increase school success among adolescents leaving in disadvantaged areas. The most disadvantaged high schools of the province are invited to develop themselves as learning organizations, to engage in a rigorous problem-solving process leading to the elaboration and implementation of action plans adjusted to the specific needs of their community. These plans have to include interventions that respect best practices principles and research-based knowledge.

The Ministry and the school boards are responsible for providing the training opportunities and technical support schools need to implement the Strategy. NANS implementation is supported by an additional funding of $25 M (CDN) per year, for five years. The NANS initiative is active all the regions of Québec, including 200 schools in disadvantaged areas. These schools, which vary greatly in size, (from 30 to 2,600 students) are located in urban, semi-urban or rural areas.

Bearing this in mind, no approach or project can have consistent results with such different implementation contexts. In order to identify the conditions conducive to explaining how different, yet also sometimes apparently similar conditions, which explains why the NANS approach is well implemented in certain schools while others do not produce the expected results, we have adopted a strategy that highlights the contrasts. We did carry out case studies in seven regions, so as to examine these contrasts.

The analyses allow for distinctions to be made among the various regions by taking into account three intrinsically linked levels of analysis, namely reports by: 1) the Ministry / regional offices 2) regional offices / school boards 3) school boards / NANS schools.
Creating intelligent accountability: questions of trust and control in Finnish and English vocational education

Mari Räkköläinen
Planning specialist
National Board of Education, Finland
Email: mari.rakkolainen@oph.fi

Kathryn Ecclestone
University of Nottingham, UK

Using the outcomes of summative assessment for national systems for evaluation and quality assurance is becoming more common in European educational systems. In Finland, the Ministry of Education requires all vocational qualifications to be attended by a practical test, demonstrating the student's attainment of the aims of vocational training. The skills demonstrations are an integral part of teaching and the learning process, desired to standardise some aspects of assessment practices whilst also diversifying student assessment and making the views of working life part of the student's vocational proficiency. In addition to promoting student assessment, the Ministry also requires the tests to improve national evaluation and quality assurance systems. The integration of national evaluation with skills demonstrations aims to abolish a system based on separate national tests and, instead, to gather the evaluation data straight from local tests flexibly administered by schools.

A development project run by the Finnish National Board of Education (FNBE) started in 2002 and the final transition to the skills demonstrations came in to force at the beginning of 2006. The development work of the new evaluation strategy at the national level will continue until the end of 2007. A feature of the Finnish development project has been to involve an external ‘peer evaluator’ to help the FNBE review progress and identify key technical, conceptual and political factors that arise in the transition to a new system. Comparison of vocational assessment and evaluation systems in Finland and the UK highlights tensions emerging in the Finnish system between local and national evaluation, the emphasis placed in national assessment on reliability versus validity, and on formative or summative purposes of assessment. These raise further questions about the desirable balance in any system between trust and control and local versus national standards, government prescription or direction and professional autonomy, and between external regulation and peer review. It is also clear that it is important to be clear about technical terms and concepts whilst recognising that they have specific cultural meanings in different systems.

Our presentation explores how these tensions have arisen in the development project and the ways in which we have tried to address them. It argues that while the principles, values and characteristics that lie behind them appear to be similar in both countries, the tensions derive from different purposes of assessment and evaluation in two very different political and cultural contexts. Peer review and evaluation in the development project reveals that comparing these tensions and reasons for them is crucial for understanding one’s own system and the pressures it faces. Our paper aims to evaluate whether it is possible to use learning results from skills tests simultaneously for student assessment and evaluating the effectiveness of training in a positive way. This may enable us to suggest ways in which we can find an acceptable balance between trust and control in the Finnish system.

Mari Räkköläinen is the leader of the national project in Finland and Kathryn Ecclestone works as an external expert and peer evaluator of the project. For further details, please contact Mari Rakkolainen.
How to develop and evaluate a web-based course on the use of national evaluation in Norwegian schools

Professor Lars Monsen
Lillehammer University College, Norway
Email: Lars.Monsen@hil.no

Bjorn Nilsen
Senior Lecturer,
Sor-Trondelag University College, Norway

In the paper we present a course developed for teachers and head teachers in primary and lower secondary schools in Norway in how they could use data from different forms of national tests and evaluations. These tests/evaluations are gathered on a website “Skoleporten.no”. When it was launched three years ago, the Minster of Education argued that this website could be used for school development and quality assurance. To follow up this claim for school development, our group for national quality assurance in Norwegian schools, got a contract to develop a webcourse showing how this new website could be used in individual schools in their internal development and quality assurance.

The paper explains how this course was developed, how it combines this large national database of quantitative data with a process of action learning in each school. Our group also made a pilot study where we wanted to find out if the aims for the course were proper and suitable for this target group. In the pilot study ca. 40 teachers/headteachers participated. They gave short evaluating comments during the course and the evaluation of the pilot study was finished with a survey. The results indicate that national tests can be used in the internal developments of a school, but that both the head teacher and several other teachers need a minimum of evaluation competence. This competence is lacking in most Norwegian schools; therefore a course like this one is very much in need.
Evaluation approaches and methodologies applied in institutional development of a quality assurance agency

Sine Frederiksen
Methodology Consultant
Danish Evaluation Institute, Denmark
Email: sif@eva.dk

Co author
Niels Matti Søndergaard
Danish Evaluation Institute, Denmark

The purpose of the paper is to demonstrate how a public institution uses evaluation approaches and methodologies to assure the quality of its own public activities and services. The paper describes the quality assurance approach of The Danish Evaluation Institute (EVA) in this respect. This approach is designed specially to fulfil EVA’s needs for input on internal institutional performance. Finally, the presentation offers a concrete example on the challenges that evaluators meet when they work with auditing and assessment of quality assurance systems. The paper offers thereby a way as to how other public organisations and institutions can develop and assess their internal quality assurance system.

EVA is an independent institution formed under the auspices of the Danish Ministry of Education. EVA’s primary task is to initiate and conduct evaluations of education - from primary school and youth education to higher education and adult education and continuing training. Thus EVA is an integral part of the educational quality assurance system in Denmark. The presenters work at EVA’s unit for methodological development and quality assurance. This unit has among other tasks that of managing EVA’s own system of internal quality assurance.

After looking briefly at the concept of quality assurance, the paper describes EVA’s quality assurance system as an example of how a public evaluation institute uses both smaller evaluations and larger transverse measures for measuring quality. The description provides an outline of how EVA uses the method of quality assurance in the constant process of improving the services of the institute. Additionally, the presentation demonstrates how follow-up procedures can be integrated as part of quality assurance systems.

Finally, the paper examines the extent to which the quality assurance system complies with standards for quality work and points to the strengths and weaknesses of the EVA-system.

Relation of paper presentation to the overall theme Evaluation in Society: Critical Connections:
The quality assurance system connects evaluation approaches and methodologies to institutional development in three main ways. First of all the system assures that all of EVA’s central activities and services comply with the objectives outlined for them. Secondly, it functions as a systematic way of gathering data for institutional development and thirdly it results in a number of performance measures on the basis of which it is possible to assess progress in strategic objectives.

Relation of paper presentation to Stream 5 Evaluating institutions, programmes and systems:
Stream 5 focuses upon critical connections between evaluation, quality and improvement of social and public services. EVA’s quality assurance system is exactly an example of how an institution that evaluates public policy and services links evaluation approaches and methodologies to quality work and its own institutional development. In that respect, the paper highlights the potential of quality assurance systems in terms of contributing to institutional development and improvement.
Multiple act criteria: fair, corresponding, democratic and symmetric criteria to satisfy all stakeholder interests?!

Mr David Kriz  
University of Mannheim, Germany  
Email: kriz@tnt.psychologie.uni-mannheim.de  

Co authors  
Manuel Völkle  
Andrés Steffanowski  

In a large-scale project titled “monitoring quality in ambulatory psychotherapy” initiated by a major German health insurance company (Techniker Krankenkasse, Hamburg) many different stakeholder groups, like patients, psychotherapists, professional interest groups and health care providers are involved and affected by the results of the program.

The focus of the intervention is the impact of a computer-based feedback system monitoring patients progress over time via validated psychometric tests. The study is a randomised two-group-pre-post-design with repeated testing over the period of therapy and a follow-up one-year after end of treatment. Data collection started in summer 2005 and will be continued until 2008. Overall 403 behavioural, depth and analytical psychotherapists from three different regions of Germany participate. Total sample size of patients will be almost 2000. It will be demonstrated how the different outcomes assessment can be combined into different multiple act criteria tailored to different stakeholder groups and to an overall evaluation of the program. Empirical research results from other program evaluations, where we already successfully used such multiple act criteria in terms of patient, external, monetary and non-monetary perspectives are demonstrated. We are confident that such an approach fulfils democratic principles of synthesizing different votes and criteria in evaluating programs and interventions.
The RBM Logical ScoreCard: a tool for the evaluation of programme logic

Geoffrey Geurts
UNESCO, France
Email: a.wright@unesco.org

Co author
Professor Alaphia Wright
Coordinator, Evaluation Capacity Development
UNESCO

Evaluations seek to answer questions on efficiency, effectiveness, relevance, sustainability, and impact with respect to institutions, programmes, etc. In practice, evaluations are usually handled within three broad sets of evaluation questions, namely: Are the right things being done, are things being done right and, are the right lessons being learnt? Interestingly, these questions are based on three fundamental assumptions. It is assumed that the right things are known, that the right ways to do the right things are known, and that the stakeholders concerned in given cases have in place conditions conducive for learning the right lessons from evaluations.

The first assumption (assuming that the right things are known) constitutes the ‘root assumption’ in any evaluation. A major requirement for useful evaluations is that of verifying the status of this root assumption, as several findings can later be explained or traced back to the said status. The status of the root assumption can be established by examining the ‘logic model’ inherent in the design of the programme concerned. This is sometimes a complicated undertaking, particularly in cases where clarity about the programme logic was not established at the design stage. There has thus been a pressing need for a simple evaluation tool for assessing programme logic in giving cases. This presentation discusses one such simple evaluation tool – the Results Based Management (RBM) Logical ScoreCard. It is based on mapping the relationships between stakeholders in a programme using the universal input-output systems model. The tool is derived from first principles, and a checklist with ten evaluation questions valid for all types of programmes, presented. An illustration of the use of the tool is discussed using a training programme as an example.
Bidding for tender: the effects of ideational factors in contractual processes

Olav A Kvitastein
Associate Professor
Norwegian School of Economics and Business Administration, Norway
Email: olav.kvitastein@nhh.no

The paper discusses the managerial and methodological challenges facing academic departments and research institutions that have a substantial part of their income from evaluations of public projects and programs. The paper questions whether common contractual arrangements can be a source of confirmation bias and suggests various improvements that can reduce unwarranted outcomes of the bidding processes. I argue that theories solely founded upon the interests of actors may overlook determinant ideational factors. Thus, I argue that the effects of ideas should complement interest as explanans for outcomes of contractual processes. Suspicion of corroborative bias may undermine the legitimacy of evaluations as an integral part of the policy process and public confidence in social science methods and thus, in the long run undermine the credibility of the social sciences. The paper questions the robustness of the sportsmanship hypothesis in economics and discusses the managerial challenges of making competition sustain the quality of research.

The paper distinguishes between short-term and long-term consequences and argues that confirmation bias due to shared ideational factors pose a hazard that may carry costs that have adverse long-term consequences. Finally, the paper discusses various institutional arrangements that may preserve the benefits of competition and protect the integrity of research.
Performance monitoring for results-based management: emergent approach or old and problematic technology?

Dr Patricia Rogers
Director
CIRCLE (Collaborative Institute for Research, Consulting and Learning in Evaluation, RMIT University (Royal Melbourne Institute of Technology), Australia
Email: Patricia.Rogers@rmit.edu.au

Since good feedback can help improve performance, performance monitoring is understandably being strongly advocated as a key component of Results-Based Management. However, good intentions are not enough. If performance monitoring is not sufficiently accurate, credible, and timely, and if these limitations are not acknowledged and addressed, it can lead to worse decision-making and performance. For many policies, programs and projects, it is difficult to produce feedback that is sufficiently accurate, credible and timely, given the range of relevant domains of performance, time-lags in achieving outcomes, the limited visibility of processes and outcomes, and the confounding causal contribution of other interventions and contextual factors. Fortunately, over the past 40 years, much has been learned about how to do performance monitoring well in these difficult circumstances, and the importance of acknowledging and addressing its limitations. Unfortunately many current guides to performance monitoring omit this history and present it as a new and unproblematic technology.

This paper sets out what we have learned about performance monitoring, and the extent to which this has been included or forgotten in guides and books. The session ends by exploring the potential role for professional evaluators and evaluation associations in making this history widely known among government and non-government organisations, and in advocating for quality standards in guides to monitoring and evaluation.
Evaluation of small programs and small-scale evaluations: intensive data analysis

Professor Lee Sechrest
Department of Psychology
The University of Arizona, USA
Email: pherman@email.arizona.edu

Small data sets can often yield highly interesting and useful information when subjected to intensive data analysis. A Canadian agency carried out a randomized study of providing insurance coverage for manual intervention for back pain in an effort to reduce absenteeism on the part of workers. The small sample of only 70 workers and the heterogeneous nature of their circumstances made evaluation difficult. Three different measures of pain were available, however, at five time points, including a six-month follow-up, and it was possible to examine data in detail at the individual level, to assess trajectories of change according to several different models, and to relate change to characteristics of individuals. This relatively intensive level of data analysis produced clear evidence for the overall effectiveness of the intervention, even though the specifics of the effects were not uniform within either group. The interesting and useful results would have been missed by any traditional analysis confined to group-level data and a single model for the expected change.
Evaluation of small programs and small-scale evaluations: resampling

Barbara Brumbach
Graduate student
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: bhagenah@email.arizona.edu

It is often desirable to evaluate programs that are small in scope, and it is also often the case that limitations on resources prohibit extensive evaluations of even large programs. Many administrators and even researchers seem to believe that small data sets merit only simplistic or even superficial analyses. In fact, data sets may be small precisely because data are scarce or hard to come by, which makes those data all the more valuable. Small data sets may, if treated carefully, yield important, even probative, information. One useful tool in the analysis of small data sets is resampling, e.g., bootstrapping, in order to produce good estimates of sampling errors. One example of successful bootstrapping analysis is provided by a very small data set involving the testing of a dietary supplement in warding off depression in pregnant women. Bootstrapped estimates gave clear evidence of the effectiveness of the supplement. A second example of bootstrapping comes from a small data set pertaining to a program involving the development of interventions to prevent hearing loss among migrant workers. A data set based on 48 cases and with 12 variables needed to be factor analyzed. The analysis was accomplished by resampling correlation matrices from a population of matrices with correlations drawn from the likely sampling distributions for each value. The initial solution based on 48 cases was compared to the distribution of solutions obtained by resampling and to the average of all the solutions.
Robustness of latent growth curve modeling to missing data problems

Professor Melinda Davis  
Research Assistant Professor  
Evaluation Group for Analysis of Data  
Department of Pediatrics  
University of Arizona Health Sciences Center, USA  
Email: mfd@email.arizona.edu

The robustness of latent growth curve modeling with varying numbers of time points has not been assessed in human growth. Are there critical time points that improve the accuracy of the estimates? Are missing end points more deleterious than missing interior points? To assess this, a 636 case growth data set with nine time points over a six year period was used. Growth curves and intercepts were calculated for the intact set, and for the same cases with a variety of patterns of missing data introduced. The results of this analysis provide information regarding the point at which parameter estimate(s) became inaccurate.
Evaluation as a catalyst: Experiences of evaluating whole of government or multi-agency programs

Julie Hind
Principal
Evolving Ways, Australia
Email: julie.hind@evolvingways.com.au

Judith Woodland
Principal
Evolving Ways, Australia

Who ensures that public policy is the best fit possible for its time, resources, and situation? This paper will argue that evaluators have a critical contribution to make to the process. The contribution is at a central policy level, through the provision of quality evidence to those responsible for policy development and funding, as well as at the policy application level through processes that enable quality improvement. The role is not as a detached outsider but as an active contributor to the process. The contribution is dynamic throughout the life of the evaluation; not merely an end product. This calls for relevant evidence to be in the public realm as it arises, not deferred until the final report, particularly in the case of longitudinal evaluation.

This is a challenging role made all the more difficult when evaluating whole-of-government or multi-agency programs in which there are multiple levels of interests and requirements. How are those many interests accommodated? How can the evaluation be of value to the many interests, particularly when bureaucracy might feel the need to moderate the process and findings because of political imperatives? How does the evaluation help ensure policy and the on-the-ground application are informed and shaped by each other?

Drawing upon five evaluations of whole-of-government or multi-agency programs in three States in Australia we share our experiences about:

- the difficulties in meeting the many diverse, and often, conflicting interests and requirements;
- the methods and processes we found most useful in ensuring robust evidence was gathered;
- the strategies that worked - and those that didn't - in ensuring that the evaluations were valuable to the many diverse interests;
- the difficulties in getting evidence into the public realm; and
- where the evaluations made a difference to policies and their applications.
The theory and practice of monitoring and evaluation in South Africa: supporting the developmental state

Indran Naidoo  
Chief Director: Governance Monitoring  
Public Service Commission of South Africa  
Email: IndranN@opsc.gov.za

The South African Public Service is in the unique position of being able to formulate, monitor and evaluate its policy in a context of recently won democracy, and at a time when the practice of monitoring and evaluation occupies centre stage in Government. Apart from the political and administrative shifts that have taken place in the past 10 years, there has been a change in the ethos of the newly inscribed public service, from a repressive and discriminatory one to one that fosters developmentalism. All of this happens within a context of globalisation, which has its own imperatives and poses difficult decisions and challenges to be made.

It is at this time that M&E initiatives of government, notably from the Public Service Commission of South Africa (PSC), which is a Constitutional body set up to investigate, monitor and evaluate public service policy and practices, needs to respond to the obligation to foster the engagement between the State and its citizenry in informing public policy. This paper will draw on the work of the PSC over the past 5 years, and demonstrate how a differentiated approach to M&E has managed to draw in cross-sectoral support for evaluation, and helped to foster transparency and accountability in Government. It will also point to the challenges around evaluation use, and discuss the difficulties around promoting learning simultaneously with accountability. Finally, it will present the Government-wide M&E System (GWMES), that is being led from the Presidency as an overarching framework for M&E in Government.

The paper will address how the PSC has broken traditional boundaries by forming social compacts in evaluation, the measures used to draw in academia, non-government organisations and organised labour, the media, donors.
Influences shaping national evaluation policies: an Irish case study

Dr Gerry McNamara
Head, School of Education Studies
Dublin City University, Ireland
Email: gerry.mcnamara@dcu.ie

Joe O'Hara
Dublin City University, Ireland

Richard Boyle
Senior Research Officer
Institute of Public Administration, Ireland

There is increased concern in the educational community regarding the role played by “transnational” bodies such as the European Commission and the OECD in the formulation and implementation of national education policies. For example, it can be argued that the adoption of the Lisbon Agenda's commitment to fostering a “knowledge economy” within the EU is resulting in synchronisation of education policies across the Union.

To some this is seen as evidence of the emergence of a “magistracy of influence” across the EU and OECD regions where networks of policy professionals and researchers shape educational discourse so as to promote a neo-liberal and globalised approach to educational policies. Whatever the truth of this contention, there is evidence emerging of a changing focus at a national educational level in the area of school evaluation, much of which seems to be influenced by the language and concepts contained in the policy documents of the transnational bodies.

The purpose of this paper is to examine the impact of EU/OECD policy ideas on the approach to educational evaluation within Ireland. The paper will analyse the emerging system of school evaluation in the context of its origins and philosophy. Particular attention will be paid to the balance struck between teacher accountability and autonomy. The paper will conclude with an examination of the importance of national contexts as mediators for the implementation of transnational policy direction. In particular it will be suggested that key contextual factors are vital to the way in which evaluation policy is worked out in practice. In the Irish context such factors include the tradition of respect for education and teachers, the corporatist approach to public sector governance that constrains “hard nosed evaluation” and the power of the teacher unions.
The experience of evaluating Structural Funds interventions and the building of an evaluation-based model of innovation in public policies - the experience of Portugal

Professor António Figueiredo
Professor and Director General
Porto Faculty of Economics and Quaternaire Portugal, Portugal
Email: afigueiredo@quaternaire.pt

Before the emergence of Structural Funds programming, the evaluation culture prevailing in Portuguese public administration was practically inexistent. The lack of accountability orientations, the remuneratory system prevailing in public administration and the weak support of research-based accumulation of knowledge concerning evaluation methodologies and supporting information systems are the main factors explaining the late coming of evaluation procedures.

The rationale of Structural Funds programming introduced evaluation practices as an intrinsic element of the programming cycle and as an obligatory procedure and not as an optional practice. Due to the relevant weight of Structural Funds interventions in the budgetary lines of public policies in Portugal, we may conclude that, through the apparatus of the EU programming, evaluation practices and methods were introduced in public management.

This particular context of introducing evaluation practices and methods in public policy systems, with no prevailing evaluation culture, raises a lot of questions about the sustainability of the procedures and mechanisms that have been established regarding the EU programming evaluation. A research question can be formulated: Can we build an evaluation-based culture of decision-making in public policies enforcing evaluation rules?

The answer to this research question is organised around the critical analysis of evaluation exercises concerning the 2000-2006 programming period. The paper exploits the experience of the author in coordinating some important evaluation studies and develops some aspects such as the quality of the evaluation procedures, the supporting information systems, the convergence/divergence between the rhythms and timings of evaluations and of decision-making processes concerning the preparation of another programming periods and the impacts of evaluation in the evolution of the mainstream orientations pf public policies.

The results are heterogenous. The paper concludes that the emergence of a sustained trajectory for an evaluation-based policy-making process has a critical factor concerning the knowledge transmission mechanisms generated by the evaluation practices. The relationships between the managing authorities of the EU Operational Interventions, the planning structures of the different Ministries and the political authorities are crucial in terms of capturing evaluation conclusions and recommandations.

In a final section, the paper analysis in-depth the consequences of the evaluation practices as inputs to generate innovating patterns in public policies, concluding that the evaluation practices introduced by the EU programming apparatus may be a source of unexpected increasing returns and learning effects in disseminating an evaluation culture among the public administration services and practitioners. Some complementary factors and conditions will be necessary to implement in order to have a feasible path for generalising evaluation-based accountability orientations.
The effect of involvement and evaluation procedures on national science foundation policy about advanced technician education

Dr Frances Lawrenz
Professor
University of Minnesota, USA
Email: lawrenz@umn.edu

The US National Science Foundation (NSF) plays an important role in developing and carrying out public policy concerning science, engineering, technology and mathematics (STEM) and STEM education. This presentation focuses on how over six years program evaluation and NSF policies have interacted to affect decisions about the development and improvement of the STEM technician workforce through the Advanced Technician Education Program (ATE).

The evaluation was a multi-million dollar effort and the program supported 75-167 projects per year for about $40 to $90 million per year. Tracing the evolution of the evaluation and the policies shows the effect of involvement by various groups on the conduct of the evaluation as well as the effect of changing levels of interaction with NSF officials. Different types of involvement by stakeholders and NSF officials resulted in varying effects on policy.

The implementation of multiple evaluation techniques over the course of the evaluation provides the opportunity to reflect on the effectiveness of these different techniques in relation to acceptance of data for policy change. For example, upon the advice of stakeholders the evaluation employed a quasi-experimental design to examine one aspect of the program. This technique was applauded at first and then criticized as being too narrow and the data were not effective in terms of policy change. Different reporting mechanisms also produced differential effects on policies. For example, the evaluation used web based interactive data tables, printed reports, face-to-face meetings with policy makers, and glass bowl sessions with the policy makers as observers as mechanisms for providing information. The type of information, the presentation style and the type of people involved all interacted with the ultimate impact of the evaluation data on policy change.

In summary this presentation will discuss effects of different levels of involvement and evaluation techniques on the development of policy.
From the evaluation of a transport policy implementation to recommendations for the future policy: the Lille Metropolis Urban Travels Plan

Dr Odile Heddebaut
Researcher
INRETS DEST, France
Email: odile.heddebaut@inrets.fr

This communication presents the evaluation process that has been sorted out by the Urban community of the Lille Metropolis (LMCU) evaluation service in order to evaluate its Urban Travels Plan (PDU). LMCU is legally obliged to elaborate a PDU and to evaluate it. The main question is how to be sure that this policy is actually implemented in the way described in the programme? This question is difficult to answer as more common is some degree of incremental implementation in which a program takes shape slowly and adaptively.

This evaluation is then firstly more focused on policy implementation rather than on decision process, involving horizontal work and different stakeholders at different scales of planning. Institutional and governance aspects, linked to the share of competences, can limit its implementation. It is a three phases evaluation process: an “inner urban community phase”, a “partners phase” and an “observation phase”. We give results of the “community phase” that measures first, the degree of appropriation of this transport policy by the different LMCU municipalities. And second, we analyse the viewpoint and commitment of the LMCU departments officers involved in its implementation.

The stake is to understand if and how this PDU is actually implemented. Then we observe if this evaluation is used in the definition of the future transport policy.
Abstract 194

Type of Presentation: Symposium (abstracts 193, 194, 195 and 196)
Stream: 1
Chair: 

Characteristics of smokers of low socioeconomic status and tobacco control policy

Ms Mei-kuang Chen

Data analyst
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona,
Psychology 312, PO Box 210068,
Tucson, AZ, USA
Postcode : 85721
Telephone : 520-621-5463
Fax : 520-621-6320
Email : kuang@email.arizona.edu

We are a group of researchers (professors, evaluators, and graduate students), called Evaluation Group for Analysis of Data led by the Professor Lee Sechrest, in the Department of Psychology, in the University of Arizona, USA. We submit a symposium called Evaluation of tobacco control efforts and public policy chaired by Lee Sechrest.

This symposium includes four presenters:

Michele Walsh.(abstract 193)
Mei-kuang Chen.(abstract 194)
Frederic Malter.(abstract 195)
Michael Menke.(abstract 196)

Smoking is particularly prevalent among smokers of low socioeconomic status, but the paramount factor is education rather than income. Most recent data suggest that about 30% of persons in the low income group are smokers, and women smoke at a lower rate than men. That is in contrast to a smoking rate of about 12% in persons with a college degree. Low income smokers smoke more heavily, and among married couples, both persons are more likely to be smokers. Low income smokers show surprisingly little inclination to smoke cheaper brands of cigarettes or to obtain cigarettes from sources providing discounts. Unemployed persons are only slightly less likely to be smokers than are those with jobs. Knowledge about the health hazards of smoking is just as widespread among low as higher socioeconomic status smokers. These low socioeconomic status smokers smoke to relieve stress and smoking is associated with at least brief periods of tranquility. Tobacco control policy needs to be directed more toward giving persons with poor prospects in life reasons to quit smoking that get beyond what they already know, which is that smoking is unhealthy and that quitting smoking would not improve their lives by much.
Archival methods in assessing tobacco control policies

Mr Frederic Malter

Research Specialist
Evaluation Group for Analysis of Data
Department of Psychology
Tobacco Education and Prevention Program Evaluation and IT Uuits
The University of Arizona
PO Box 210462
Tucson, AZ, USA
85721

Telephone : 520-318-7259 ext 147
Fax : 520-318-7252
Email : fmalter@email.arizona.edu

We are a group of researchers (professors, evaluators, and graduate students), called Evaluation Group for Analysis of Data led by the Professor Lee Sechrest, in the Department of Psychology, in the University of Arizona, USA. We submit a symposium called Evaluation of tobacco control efforts and public policy chaired by Lee Sechrest.

This symposium includes four presenters:

Michele Walsh.(abstract 193)
Mei-kuang Chen.(abstract 194)
Frederic Malter.(abstract 195)
Michael Menke.(abstract 196)

Many types of public and private records are useful in evaluating and guiding public policies in a wide range of areas. Tobacco control policies are often in the form of tax increases on tobacco products. When tax increases are enacted, there is likely to be a corresponding change in the sale of tax stamps that are used to legitimate the sale of tobacco products. Data on sales of tobacco tax stamps in Arizona and other states show the effects of tax increases by a temporary drop in sales, sometimes preceded by a temporary increase in sales of tax stamps as merchants stock up on the cheaper stamps. Subsequent to the temporary decrease in sales, the data generally show a recovery but to a lower level and indicate a generally decline in sales of cigarettes. Data from other sources, however, e.g., survey responses, indicate that increases in taxes may have rather little effect on prevalence of cigarette smoking but may contribute to a reduction in amount of smoking.
Abstract 196

<table>
<thead>
<tr>
<th>Type of Presentation:</th>
<th>Symposium (abstracts 193, 194, 195 and 196)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream:</td>
<td>1</td>
</tr>
<tr>
<td>Chair:</td>
<td></td>
</tr>
</tbody>
</table>

Air quality monitoring and policies involving smoking bans

Mr J. Michael Menke

Health Research Associate
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona,
Department of Psychology,
Psychology 312, PO Box 210068,
Tucson, AZ, USA
85721

Telephone : 520-621-5463
Fax : 520-621-6320
Email : menke@email.arizona.edu

We are a group of researchers (professors, evaluators, and graduate students), called Evaluation Group for Analysis of Data led by the Professor Lee Sechrest, in the Department of Psychology, in the University of Arizona, USA. We submit a symposium called Evaluation of tobacco control efforts and public policy chaired by Lee Sechrest.

This symposium includes four presenters:

Michele Walsh.(abstract 193)
Mei-kuang Chen.(abstract 194)
Frederic Malter.(abstract 195)
Michael Menke.(abstract 196)

In many local communities in the United States, smoking has been banned in various public and quasi-public places such as worksites, restaurants and bars, and recreational facilities. The principal aim of such bans is to improve air quality in public places. Two questions that immediately arise are whether there will be compliance with the smoking bans and by how much air quality will improve. Current technology for measuring air quality makes it possible to measure the density of very small particles of minerals resulting from smoking and that are believed to be critical in the development of breathing problems and lung cancer as a result of smoke exposure. A prospective study of air quality in restaurants and bars in an Arizona community before and after the enactment of a ban on smoking showed generally good compliance and a very large improvement in air quality. The air quality data were supplemented by observational data concerning the numbers of patrons, numbers of smokers, and so on. Compliance was not perfect in some establishments because of a single special occasion that apparently attracted smokers either not aware of the ban or unwilling to abide by it. That occasion provided good evidence for the sensitivity of the air quality monitoring procedure. These data are consistent with the few other air quality monitoring studies and with observational data from other locales.
Evaluation of tobacco control efforts and public policy

Surveillance methods in evaluation: survey methods

Dr Michele Walsh
Senior Evaluator
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: mwalsh@email.arizona.edu

Surveys of one sort or another have been the customary ways of assessing the effects of social policies that are designed to produce incremental changes over time. A notable instance is tobacco control, almost universally assessed by household surveys, in recent years by telephone. Surveys, perhaps especially those by phone, are becoming more difficult and perhaps less trustworthy. In the United States cell phones are not reachable by random digit dialing, and completion rates are falling even when connections are made. Respondents tend to be female, and households with lower levels of income and education are under-represented. Cultural and language minorities are difficult to reach and to interview. The limitations of surveys are illustrated by data from large scale surveys carried out in the United States regarding prevalence and incidence of smoking and effectiveness of prevention and control efforts. More extensive use of sophisticated multivariate analysis can help to overcome some of the limitations of surveys and extend their usefulness. Multivariate analyses give a clearer picture of subsets of the population and the factors influencing their smoking and initiation into smoking.

Characteristics of smokers of low socioeconomic status and tobacco control policy

Mei-kuang Chen
Data analyst
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: kuang@email.arizona.edu

Smoking is particularly prevalent among smokers of low socioeconomic status, but the paramount factor is education rather than income. Most recent data suggest that about 30% of persons in the low income group are smokers, and women smoke at a lower rate than men. That is in contrast to a smoking rate of about 12% in persons with a college degree. Low income smokers smoke more heavily, and among married couples, both persons are more likely to be smokers. Low income smokers show surprisingly little inclination to smoke cheaper brands of cigarettes or to obtain cigarettes from sources providing discounts. Unemployed persons are only slightly less likely to be smokers than are those with jobs. Knowledge about the health hazards of smoking is just as widespread among low as higher socioeconomic status smokers. These low socioeconomic status smokers smoke to relieve stress and smoking is associated with at least brief periods of tranquility. Tobacco control policy needs to be directed more toward giving persons with poor prospects in life reasons to quit smoking that get beyond what they already know, which is that smoking is unhealthy and that quitting smoking would not improve their lives by much.
Archival methods in assessing tobacco control policies

Frederic Malter
Research Specialist
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: fmalter@email.arizona.edu

Many types of public and private records are useful in evaluating and guiding public policies in a wide range of areas. Tobacco control policies are often in the form of tax increases on tobacco products. When tax increases are enacted, there is likely to be a corresponding change in the sale of tax stamps that are used to legitimate the sale of tobacco products. Data on sales of tobacco tax stamps in Arizona and other states show the effects of tax increases by a temporary drop in sales, sometimes preceded by a temporary increase in sales of tax stamps as merchants stock up on the cheaper stamps. Subsequent to the temporary decrease in sales, the data generally show a recovery but to a lower level and indicate a generally decline in sales of cigarettes. Data from other sources, however, e.g., survey responses, indicate that increases in taxes may have rather little effect on prevalence of cigarette smoking but may contribute to a reduction in amount of smoking.

Air quality monitoring and policies involving smoking bans

J. Michael Menke
Health Research Associate
Evaluation Group for Analysis of Data
Department of Psychology
The University of Arizona, USA
Email: menke@email.arizona.edu

In many local communities in the United States, smoking has been banned in various public and quasi-public places such as worksites, restaurants and bars, and recreational facilities. The principal aim of such bans is to improve air quality in public places. Two questions that immediately arise are whether there will be compliance with the smoking bans and by how much air quality will improve. Current technology for measuring air quality makes it possible to measure the density of very small particles of minerals resulting from smoking and that are believed to be critical in the development of breathing problems and lung cancer as a result of smoke exposure. A prospective study of air quality in restaurants and bars in an Arizona community before and after the enactment of a ban on smoking showed generally good compliance and a very large improvement in air quality. The air quality data were supplemented by observational data concerning the numbers of patrons, numbers of smokers, and so on. Compliance was not perfect in some establishments because of a single special occasion that apparently attracted smokers either not aware of the ban or unwilling to abide by it. That occasion provided good evidence for the sensitivity of the air quality monitoring procedure. These data are consistent with the few other air quality monitoring studies and with observational data from other locales.
Shifting Paradigms, Organic Logic in Value for Money Auditing

Dr Rudi Turksema
Audit expert
Netherlands Court of Audit, The Netherlands
Email: r.turksema@rekenkamer.nl

Coming from a financial audit (accountancy) background most audit offices carry out Value for Money (VFM) audits with research concepts adopted from financial audits. The traditional focus emphasizes compliance with formal procedures, organisational provisions, and accountability. Common questions in this paradigm are: is policy based on valid information, is information on policy implementation and policy results available, are responsibilities clearly defined?

From a control or ex ante perspective on policy these questions are relevant, but the answers to such questions do not lead to insights in the substantial results and outcomes of policy. Also, the traditional focus does not provide plausible explanations for unsatisfactory policy results, other than those that emphasize administrative and procedural shortcomings. Lastly, the traditional focus is mostly limited to those explanations that lie within the sphere of influence of the audited body. The focus on procedures may even obscure the insight into policy substance and effects. This could lead to strategic behaviour on the part of the audited body. Therefore, we propose to shift towards a paradigm that augments the traditional procedural logic in VFM audit with a logic that emphasizes the substantial quality of policy and policy outcomes. We call this organic logic.

VFM audits in the organic logic paradigm take the social problem at which policies are directed as a starting point. What is the problem in society and does policy actually contribute to its solution? If problems are not solved by policy, the auditor has to find explanations for discrepancies between policy, policy results and the (persistent) social problem. It is at this point that the organic logic comes in. We propose that auditors also investigate other possible explanations than the organisational provisions. These explanations may range from an inappropriate policy theory or lacking support among stakeholders. Auditing from an organic logic paradigms implies that auditors broaden their methodological toolkit. Not only interviews and literature surveys, but also comparative case studies, regression analyses and other methods that focus on policy substance, should be part of their toolkit. In this paper we present an outline of what an organic approach to VFM audit may look like in practice.
The Usage of PART in the European context - possibilities and caveats

Alfred Radauer
Senior Researcher
Austrian Institute for SME Research; Austria
Email: a.radauer@kmuforschung.ac.at

Co author
Klaus Zinoecker
WWTF, Austria

In the past years, and with the increasing number of government support programmes implemented, there has been a growing demand for performance measurement in many countries. In particular, many policy makers call for a thorough assessment of all support programmes in a way that allows for easy comparisons of the effectiveness of different schemes. The way and extent to which this is done varies considerably across countries, depending on the respective evaluation culture.

In the U.S., reporting under GPRA (Government Performance and Results Act) and/or "parting" of support programmes applying the PART (Programme Assessment Rating Tool) questionnaire is mandatory for all government funded support programmes, regardless of policy fields. PART is of special interest, as it uses a standardised set of questions and a numerical grading system which should allegedly allow for direct comparisons with regard to effectiveness. Such an over-arching instrument does not yet exist in Europe.

The aim of the underlying paper is to discuss benefits and caveats of implementing PART features in Europe and, in particular, in Austria. The paper draws on a literature analysis and a quantitative analysis of PART scores.

In our paper we will argue that:

- PART is discussed controversially in the U.S. The number of identified benefits is significantly lower than the number of so-called "challenges".
- PART scores are only weakly correlated to funding decisions.
- The European and Austrian evaluation system are less quantitatively oriented and less output-figure driven. Perceived advantages in better catching the context of support programmes are counteracted by lower comparability of evaluation results.
- Implementing PART 1:1 in Austria or Europe does not seem feasible. However, some features of PART (e.g. defining mandatory elements for programme evaluations) could improve the evaluation system, if certain provisions are being made (e.g., the commissioning of regular context studies).
The use of Social Network Analysis in regional policy evaluations

Richard Hummelbrunner
Consultant
ÖAR Regionalberatung, Austria
Email: hummelbrunner@oear.at

Regional policy is characterized by the interaction of economic, social, cultural and physical resources within a territory. The success of programmes or projects depends on the quality of collaboration, notably communication and behaviour patterns, between key actors having access to or being responsible for these resources. Therefore these aspects need to be taken into account in evaluation and a more thorough understanding of relationships allows to identify factors which are critical in explaining present or past performance as well as to suggest changes in the future.

Social Network Analysis (SNA) is a tool which lends itself well in this direction. It is a method to understand networks and their participants, that is applied in wide range of disciplines and which offers a variety of techniques for measuring, visualising and simulating relations in social networks or systems. It provides both a visual and a mathematical analysis of relationships and flows between people, groups, organizations etc. An important feature of SNA are measures which allow to assess the importance of individual actors (nodes) via their location in the network (e.g. centrality, network reach, boundary spanners).

At the beginning, the paper outlines the need and rationale for the use of SNA in evaluating regional policy. The core of the paper is the description of two cases where the author has used SNA in evaluation assignments:

- The evaluation of development structures in an Austrian province, where SNA was used to assess inter-organisational relations within and beyond the territory under study
- The evaluation of cross-border co-operation, where SNA was employed to analyse on the relationships among actors in selected fields, and of cross-border projects which have been funded under the INTERREG Community Initiative.

The final section of the paper summarizes the experience gained and explores further potential uses of SNA in regional policy evaluations.
Evaluation and its relationship to environmental policy, environmental actors/sectors and environmental systems

Joos Gysen
Catholic University of Leuven, Belgium
Email: joos.gysen@hiva.kuleuven.be

This paper addresses the place of evaluation in environmental policy. Firstly it levels the playing field by describing the particularities of the environmental system and of environmental policy. Secondly it brings a methodological perspective to this playing field by connecting environmental evaluation needs to evaluation methods taking the playing field into account. Thirdly it addresses the emergence of new sectors due to environmental policy. These sectors respond differently to evaluation. In this part we share our experience with evaluations in 4 new sectors, due to environmental policy, and how the interaction with these sectors influences methodological choices. Lastly the paper gives an overview of evaluation methods that both fit environmental policy and addresses the new evolutions in evaluations methods in general.
Evolution of evaluation thinking in Kenyan health sector

Kamau Njoroge
Disease Surveillance
Medical Emergency Relief International (MERLIN), Kenya
Email: jonseenjoroge@yahoo.com

Introduction
Kenya's Ministry of Health (MoH) has been in existence since 1920's. Its responsibility is delivery of health care to the nation. There has been tremendous growth of the health sector since independence in 1963. From late 1980's, donor and public pressure has grown pushing MOH to match outputs with resources.

To cope with the pressure MoH formed Health Management Information Systems division at the national level to document its impact in the community. Routine reports are collected by facilities and transmitted to the district level where compilation is done and transmitted to national level. The objective was to help the national office monitor the progress on the ground and to a lesser extent help in planning as part of decentralization. This strategy started in 1983 with introduction Rural Focus for Development strategy. Districts were to manage their own affairs by planning and evaluating their activities. However, many problems abide.

Weaknesses
· Lack of M&E policy in MoH
· Duplication of data collection tools
· Disregard for data in intervention
· Planning without implementation and implementation without planning
· Lack of supportive infrastructure
· Unreliable data in remote districts due to lack of trained health workers

Opportunities
· Existing network of health facilities
· Well trained work force

Progress so far
Some divisions have come up with M&E policy e.g. TB and Leprosy

Way forward
· MoH to design M&E policy from which divisions can borrow and tailor according to their needs.
· Revision and merging of duplicated tools;
· Medical schools to start offering M&E as a discipline
· District Health Management Team (DHMT) to be involved in designing of tools at the national level
"Ecological" issues in the evaluation of the experimental short school week in Israel

Dr Nir Ressissi
Oranim Academic College of Teacher Education, Israel
Email: nirrs@magal.org.il

Varda Gil
Ministry of Education, Israel
Email: vardagil2002@gmail.com

At the end of 2003 the Ministry of Education published a call for an evaluation study of the short school week established in most regional education systems in Israel. The decision to implement the short school week had been taken during the summer of 2003, and its experimental format softened the opposition voiced by the teachers' organizations. The researchers answering the call were asked to plan two research stages: the first small in scope, focusing primarily on the educational-behaviorist aspect; and the second large in scope, focusing on the economic aspect. A large sum of money had been allotted for this purpose, and we decided to meet the challenge.

In March 2004 we were informed by the Research Committee, which included representatives of the interested bodies and was headed by a professor of education, that our proposal had been accepted over the others (despite our position of inferiority as a college vis-à-vis the universities). At the same time, the acceptance included two caveats - administrative and methodological - which meant we had to make do with implementing the first, smaller, part of the study only. Later, we discovered that this was due to the fact that the Ministry of Education was starting at the time to get organized towards the implementation of a short school week in the entire educational system.

Our updated proposal was approved only in May 2004, at which time the report of the "National Taskforce for the Improvement of Education in Israel" were also made public, including a recommendation which suited the Ministry's purposes: "All kindergartens and schools will operate five days a week only". Thus, we found ourselves beginning the evaluation study itself with the entire educational system in an uproar. In fact, the aforementioned recommendation received much public attention, which affected the relevance of our study not only for the participants but also for its initiators, who put pressure on us to bring out an unplanned interim report in the summer of 2005 and publish the final report in the winter of 2005-6.

The lecture presents three circles of "political" influence around the evaluation study, and focuses on the manner in which the (overt and covert) positions of the involved bodies were expressed (directly or indirectly): the inner circle - positions expressed by academic members of the Research Committee; the middle circle - positions expressed by the Ministry of Education representatives; the outer circle - positions expressed by the public and the political system. In our opinion, all these influenced the nature and essence of the evaluation report, which might have been much more valuable in a different "ecological" context.
Finnish wonder and Danish trauma: Impacts and paradoxes in the International assessment trend

Dr Oscar Oquist
Director of Education
Swedish National Agency for Education, Sweden
Email: oscar.oquist@skolverket.se

Sten Soderberg
Swedish National Agency for Education, Sweden

In this article I raise some questions about the strong impact international comparative assessments has in the field of education and curriculum with the Nordic countries as an example. Strong structural adjustments in the Danish and Norwegian educational system have now been initiated as a direct answer to the latest PISA and PIRLS results. Finland who has came out very good in the international comparisons has on the other hand been praised for their high educational quality and looked upon as a model for the other countries. The Nordic examples show that we in different national contexts use the international studies in different ways according to different history and different socio-economic positions. In the article I raise some questions about the longrun impact of the international educational hegemony and ask if this ranking of countries unwittingly leads to a globalisation effect or drift to the mean where national curricular identities successively are erased.

Another point I will discuss in the paper is about democracy and selfsteering and the problematic double-bind new and small countries as Norway or Estonia are put into by the international comparisons. The question is how they in their curricular frameworks can sustain their own history important for cohesion, meaning-making and nationbuilding and in the same time adjust to the global trends towards privatisation, deregulation and standardisation and a “systeme mondial” in the educational arena where the theory of literacy of the PISA-study have come to play an important role.
Effectiveness analysis of economic instruments in resource policy

David Legg
Environment Agency for England and Wales, UK
Email: David.Legg@environment-agency.gov.uk

Mette Skovgaard
European Topic Centre on Resource and Waste Management, Denmark

Co authors:
Henry Leveson-Gower
Environment Agency for England and Wales
Massimiliano Mazzanti, Susanna Paleari and Roberto Zoboli
Institute for Economic Research on Firms and Growth, CERIS CNR Milan
Marton Herczeg
The Regional Environment Center for Central and Eastern Europe
Raimund Bleischwitz
Wuppertal Institute
Gorm Dige
European Environment Agency

A key developing area of policy focus for the EU is the efficient use of resources as initiated by the EU ‘Thematic Strategy on the Sustainable Use of Natural Resources’ and the prevention and recycling/recovery of waste as defined by the ‘Thematic Strategy on Waste Prevention and Recycling’. In both strategies, specific attention is given to economic and market based instruments, e.g. taxes on waste, as useful tools for improving environmental policy performance. The 6EAP includes the idea of ‘pushing the market to work for the environment’ by both including environmental costs in the price of products and services, and developing agreements between environmental policy institutions and economic actors.

The objectives of the study are to:

- Increase understanding of the efficient and effective use of economic instruments in resource management in EEA member countries (i.e. to share good practice and increase knowledge and understanding in this area)
- Develop a framework for evaluating ex post the efficiency and effectiveness of economic instruments. In particular, to find evidence on how the inclusion of economic instruments in complex policy packages can improve their overall performance and how economic instruments can be used as a prevention tool in terms of resource use and waste generation.
- Undertake an ex post evaluation of an aggregates tax (or equivalent) in four different countries using the above-mentioned methodology. The aim is to compare approaches and evaluate the use of aggregate taxes to inform and influence policy development in other countries.

The countries that have been selected include the Czech Republic, Italy, Sweden and the UK so that a range of experiences (e.g. geographical, political etc.) can be included in the study.

For the conference in the applied methodology and the draft results will be presented. The study is to be finalised in May 2007.
Peer review and evaluation in the new EU governance: lessons from the OECD peer reviews

Markku Lehtonen
Ph.D., Research Fellow
Sussex Energy Group
SPRU – Science and Technology Policy Research
University of Sussex, UK
Email: M.Lehtonen@sussex.ac.uk.

As a result of the mounting difficulties in implementing policies in the traditional top-down manner, the European Union is increasingly relying on new, network-led modes of governance. The new EU approach, as described in the 2001 White Paper on governance, is characterised by attempts to increase public participation in EU decision-making and to introduce new instruments that would allow greater flexibility at national level. Within this new orientation, the so-called Open Method of Coordination (OMC) is a case in point, essentially consisting of policy guidelines and goals set up at EU level, indicators for benchmarking national performance, and periodic monitoring, evaluation and peer review of member states. With its recognition of the subsidiarity principle, focus on collective goals and discipline, as well as the objectives of policy coordination, learning and improvement of performance, the method bears striking resemblance to the “peer review” method applied within the OECD since the 1960s.

Looking specifically at environmental policy and sustainable development, this paper seeks to identify the lessons that EU governance can draw from the experience of two OECD peer reviews – the Environmental Performance Reviews (EPRs) and the inclusion of a sustainable development section in a full cycle of the Economic Surveys in 2001-2004. In particular, the paper examines the success of the reviews in improving policy coordination and integrating environmental and sustainability concerns into sectoral policies. The analysis seeks to explain reasons for the failure of the Economic Surveys, despite their high political status, to generate more than very modest impacts, and contrasts this experience with that from the EPRs, which are politically less visible, but whose approach is considerably closer to the principles of environmental policy integration. The OECD experience highlights a number of issues that merit more attention when analysing the role of evaluation and peer review in the new EU governance: the importance of political leadership, the difficulties in achieving consensus among a heterogeneous group of OECD member countries; the role of the dominant ‘organisational discourse’ in shaping the influence of reviews conducted by an international organisation; the need to motivate key ‘change agents’ within the national administrations; and the importance of country-specific framework conditions in shaping the influence of peer reviews.
Evaluating business investment incentives co-funded by the ERDF: methods and empirical evidence from the Obj.2 areas of Italy

Professor Daniele Bondonio
Associate Professor
Università del Piemonte Orientale /Research Center for Evaluation Studies in the Public Sector, Italy
Email: daniele.bondonio@sp.unipmn.it

Co author
Robert T. Greenbaum
The Ohio State University, USA
Email: greenbaum.3@osu.edu

Over the past decade business investment incentives co-funded by the European Regional Development Fund (ERDF) have been offered in more than 80 Objective 2 areas covering 18% of the EU population [with approximately 11% of the entire EU budget for the fulfillment of economic and social cohesion objectives devoted to such incentive packages]. Despite the wide popularity of these initiatives, very little reliable ex-post evidence of their employment impact is yet available to help EU policy makers refine future geographically targeted economic development policies. To date, employment impact results are derived by large through two evaluation practices: application of standard macroeconomic multipliers to the volume of investments co-financed by the ERDF and solicitation of entrepreneurs' judgments on the effectiveness of the incentives in affecting their investment behavior. Both types of procedures have serious drawbacks. Conducting reliable ex-post impact evaluations based on actual pre-post intervention data is very demanding as it requires methods to disentangle changes due to the incentive programs from changes due to all of the economic and social factors exogenous to the program interventions. Italy presents an ideal opportunity to evaluate the impact of Obj.2 business investment incentives because of the unique availability of data regarding both the program incentives paid to each assisted SME and the firms' yearly employment changes recorded by the Italian Social Security Agency.

Based on such data, this paper develop a parametric difference in difference evaluation model that estimates the employment impact of the Obj.2 business incentives net of all changes due to the economic trends that are exogenous to the program intervention. The model also allow the marginal employment impact of the programs' financial generosity to be estimated along with differences in the employment impact due to different degrees of pre-intervention industrial decline in the treated units. Results of the analysis show a positive and significant marginal employment impact in the assisted areas when the financial generosity of the incentives is increased. The estimated employment impacts, however, are lower than those offered by the evaluation reports that either apply standard macroeconomics multipliers to the volume of subsidized investments or collect entrepreneurs' judgments on the employment effectiveness of the program. Accounting for the employment outcomes by degree of pre-intervention industrial decline, the analysis further finds that the incentives were most effective when targeting production in province-industry pairs that had the least severe declines during the years prior to the program intervention.
Evaluating European Union Programs: an integrative methodology

Inmaculada Serra Yoldi
Departamento de Sociología
Universidad de Valencia
E-mail: Inmaculada.Serra@uv.es

Co author
Elvira del Pozo Aviñó

The methodology we present in this paper was utilized in the evaluation of an European Union’s Equal Program. For this evaluation we employed a methodological design that is intricate, given the complexity of the Project’s goals, its longitudinal and diachronic character, and the diversity of the organizations and agents participating in its development. An internal evaluation was conducted and analyzed from two different perspectives: a quantitative and a qualitative one. With regard to the quantitative perspective, the evaluation’s objective was two-fold: on one hand, we conducted a process evaluation, and on the other hand, we pursued an outcome evaluation. In order to achieve coherence and consistency to the whole of the evaluating process, these two dimensions were kept interdependent and feed one another. In relation to the qualitative perspective of this evaluation, the primary objective was to carry out an in-depth study of the opinions and values of the different stakeholders, both those responsible for providing the services and the recipients of such services. The quantitative aspect of the evaluation gathers data following several objective indicators measuring the degree to which the program is operating. To obtain information on the qualitative process of the evaluation, we utilized several research techniques, such as in-depth interviews, group discussions and the content of analysis directly drawn from the discourses of those interviewed. Combining both perspectives during the evaluation analysis has offered articulating and converging different paradigms resulting into an integrative methodology. The purpose of this methodology was three-fold: first, to enrich the research, second, to explore new aspects of the topic under study, and third, to diminish the limitations encountered when using only one method by adding the strengths provided by the use of other methods.
Synthesis and reflections: the future of international evaluation

No abstract available
Improving performance? The challenges to evaluation as a management tool

Dr Steffen Bohni Nielsen  
Manager  
Rambøll Management, Denmark  
Email: sni@r-m.com

Nicolaj Ejler  
Rambøll Management, Denmark

Underlying the connection between evaluation, quality and improvement is the critical question: how can we establish a relevant format of evaluations so it can drive improvement? - From the perspective of many public managers and field workers evaluation reports often produce a body of knowledge that appears too late and too much (to bother reading) to be useful as a management tool. This poses a significant problem for external and internal evaluators alike. - How can we structure evaluations as a management tool in a form that renders the information relevant, timely, and useful in the very implementation of public interventions? Drawing on experiences from evaluating and advising on social policy, health and labour market interventions the presenter argues that the very format of evaluations must be rethought and that the boundaries between evaluation and performance management may need to be redefined in the process of utilizing evaluation as a management tool. Under three headings the presenter investigates the challenges posed to evaluation.

**Design:** In order to make evaluation a useful tool for improvement of public interventions the production of evaluative knowledge must be carefully tied to the mission and targets of the intervention. In doing so, processes such as defining the theory of change, developing meaningful output and outcome indicators, and the requirements of a M&E system become subject to negotiation between the evaluator and the evaluated. The presenter argues that integrating evaluation intimately in the policy design process is crucial for making full use of an evaluation. Drawing from the theory of change tool policy design, evaluation and performance management can be closely interlinked. In this regard, evaluation transgresses the boundaries to performance management and measurement.

**Implementation:** For many public managers and field workers the evaluation process is a cumbersome affair that takes time and resources away from their work. The data produced play a minimal role in their service delivery. A crucial challenge is to make the production of data a meaningful task to those producing the data.

In the presentation we explore in ways that secure management and field workers buy in to the process and create incentives to make utilization of the evaluative knowledge produced. In fact, the presenter, argues, integrating this knowledge production to the work processes increases the utility of evaluation. Again, this exercise transgresses the boundaries to performance management.

**Meeting:** A pivotal process in performance management is the institutionalization of the situations in which management and field workers engage in discussions and interpretations of the output and outcome data produced. In short, regularized meetings where data are utilized to improve quality and results. Drawing on examples from local and national government performance management projects the presenter reflects on the constituent parts of the meeting.

In sum, the presenter argues that the form of evaluation is inadequate as a day-to-day managerial tool. If the utilization of evaluations should be enhanced this implies that theoretical and practical boundaries to performance management and measurements must be redefined.
Towards a qualitative approach in the evaluation of rural
development policies

Beatriz Izquierdo
Professor assistant in Sociology
University of Burgos, Spain
Email: bizquierdo@ubu.es

Most of what is known about rural development policy is obtained from the evaluations carried out by the Member States (some of the regions) and the Commission. These exercises provide a very useful information in order to acknowledge what is going on in the program: number of projects financed, number and type of beneficiaries, money invested etc. The data obtained is normally divulged in quantitative terms, in order to answer the questions and objectives contained in the evaluation policies. This mechanical model, however, omits that programs are developed at local contexts and are mediated by local agents, and what is more important, that evaluation must be adapted to the socio-economic and cultural context. Besides, their evaluations offer limited information on the ‘field’: its participants and beneficiaries (public and private ones), and their perceptions and opinion about rural development policies.

The paper highlights the incorporation of the opinion of the beneficiaries to the evaluation of rural development policies. The paper shows the main results from that experience through several in deep interviews conducted to different beneficiaries (enterprises, local development agencies, local mayors, farmers etc.) from rural areas of the Basque Country.
A trident approach to programme evaluation

The symposium will present the details of an evaluation methodology based upon the stated outcomes of a project, the process(es) by which the project was delivered and the elicitation of multiple stakeholder perspectives to identify what those associated with the project actually thought about it. The first paper will outline the methodology of the Evaluation Trident and the way it is applicable to a wide range, this will be followed by 5 papers, outlining the varied policy and practice areas that have been evaluated using this methodology.

A Trident Approach to Programme Evaluation

Professor Roger Ellis OBE
Director of Research Policy
University of Chester, UK

Dr Elaine Hogard
Leader of the Social & Health Evaluation Unit
University of Chester, UK

This symposium describes a trident approach to programme evaluation developed by the Social and Health Evaluation Unit at the University of Chester and includes three case studies of the method in use. The first paper describes the trident approach and its methodological implications. The trident focuses attention on three main aspects of a programme, namely its outcomes, its process and stakeholder perspectives on it. Three basic questions are therefore posed regarding the programme: did it work? what happened? and what did people think of it? Evaluation is essentially posing research questions and gathering data to answer them. Questions, data gathering methods and issues are discussed for each prong of the trident. Addressing the questions identified requires a mixed method approach involving quantitative and qualitative data and both positivistic and phenomenological orientations. For example outcomes must often be measured in publicly accessible and aspirantly objective forms whereas stakeholder perspectives are, of their mature, phenomenological. The study of process is identified as a relatively neglected area and its significance for the elucidation of professional/service user interaction is stressed. The utility of the three areas of evidence is discussed in relation to funders, policy makers and providers. The partnership between contractors and evaluators is conceived as a negotiation to determine the precise questions and data gathering which will be appropriate for each prong of the trident. At the beginning a crucial issue is the clarification of the objectives of the programme and the outcome measures which will provide evidence of the extent to which the programme has met its objectives. Second, consideration is given to the process by which the programme is delivered and the extent to which this will have to be explored as a novel activity or checked for compliance to predetermined standards. Thirdly the multiple stakeholders affected by the programme, including centrally service users and providers must be identified and methods agreed to solicit their opinions.
Evaluation of the ‘Safer Homes’ domestic burglary reduction initiative

Jeremy Warren
Research Officer
Social & Health Evaluation Unit
University of Chester, UK

Introduction
The ‘Safer Homes’ project was implemented by a police force in the North West area of England as a force-wide response to poor performance in respect of domestic burglary. Members of the same force had recently written the National Good Practice and Tactical Options Guide for Domestic Burglary; and it was this document that had formed the backbone of the project. The evaluation was conducted using the Ellis & Hogard ‘Trident’ model, identifying the key outcomes for the project, the processes used to reach those outcomes and identification of stakeholder views on the degree of success reached by the project.

Outcomes
In consultation with the project implementation team, the evaluators identified four main outcomes against which the project could be fairly judged:
1. Delivery of a better quality of service to victims of domestic burglary
2. A reduction in the overall number of domestic burglary incidents across Cheshire
3. A higher quality of investigation leading to a higher domestic burglary detection rate
4. An improvement in the relationships between crime reduction agencies in Cheshire
5. The fifth key objective was defined by the Home Office Police Standards Unit.
6. Evaluation of the role played by the Force Burglary Co-ordinator

The results of the evaluation were positive for all outcomes, with the exception of number 4, where it was found that the project had had little effect on improving communication at the strategic partnership level.

Process
The force undertook a ‘root and branch’ appraisal of the way in which they dealt with domestic burglary cases, implementing where appropriate the good practice that had been gained from other forces. Where good practice had not been found the implementation team took the opportunity to revise or remove outdated policies and procedures.

Multiple Stakeholder Perspectives
Three main groups of stakeholders were identified during the course of the evaluation:
- Constabulary staff involved with the domestic burglary process
- Victims of domestic burglary
- Partnership agencies with an interest in burglary reduction.

From interviews conducted with members of all three groups it was found that the changes introduced to the domestic burglary process through the ‘Safer Homes’ project had had a significant positive impact on the way in which the police dealt with domestic burglary. This was reinforced by the victims interviewed, many of whom felt that the service received was

Recommendations for Practice
The evaluation project produced several service level recommendations for practice as well as an implementation checklist for the adoption of good practice, irrespective of the policy area concerned.
Evaluation of the impact of Lead Nurses within a UK hospital

Elizabeth Gibbs
Research Assistant
Social & Health Evaluation Unit
University of Chester, UK

Introduction
Lead Nursing was introduced on the Child and Adolescent Unit at Leighton Hospital by the Practice Development and Research Unit (PDRU) in 2003 as part of an alternative system of care management, encouraging clinical leadership. Fourteen Diagnosis Related Group (DRG) teams were created and allocated a Lead nurse. The Evaluation Trident (Ellis & Hogard, 2005, 2003) was employed to evaluate Lead nursing on the unit through the exploration of outcomes, processes and multiple stakeholder perspectives. A qualitative research methodology was employed which incorporated interviews with Lead and other staff nurses on the wards.

Outcomes
There was no specific outcome outlined for Lead nursing, thus in order to measure the outcomes of the lead nursing system the extent to which the Lead nurse and DRG team responsibilities had been fulfilled was investigated. Evidence was found that indicated only half of the lead nurse responsibilities and one of the four DRG team responsibilities are currently undertaken.

Process
The proposed and current working practices of the Lead nursing system were mapped, revealing that Lead nurses in several DRG areas are keeping up to date with developments in care; and providing feedback to staff nurses via informal teaching sessions and updating protocols and guidelines. The process analysis indicated that Lead nurses have not been using all available methods to communicate updates in practice and that the DRG team setup is not currently functioning effectively.

Multiple Stakeholder Perspectives
Feedback from the Lead and other staff nurses working on the Child and Adolescent unit indicated that the Lead nursing system is working particularly well in three diagnosis areas; namely diabetes, oncology and cystic fibrosis and that staff also believe speciality contacts to be a beneficial addition to the unit. The low profile of Lead nursing amongst staff members however, was highlighted and time restraints were seen to be a challenge to the full functioning of the system on the ward.

Recommendations for Practice
Recommendations were made to address the areas where responsibilities were not being adequately fulfilled, to widen the participation of the DRG teams and also to raise the profile of Lead nursing on the Child and Adolescent unit.
Evaluation of the ‘Shaping the Future’ Education and Training Project

Jackie Ellis
Research Assistant
Social & Health Evaluation Unit
University of Chester, UK

Introduction
The Shaping the Future for Primary Care Education and Training Project was a collaborative partnership between the North West Universities Association and the North West Development Agency and seven North West Universities. The project also brought together for the first time all key partners health, social care and education sectors who are involved in supporting the delivery of health and social care in the North West Region.

One of the main aims of the project was to encourage collaboration between Higher Education Institutions, through a research project which had potential value for a number of key stakeholders within the region involved in the delivery and development of integrated health and social care sectors within the primary care sector.

Project evaluation followed a two-fold approach, with part of the evaluation undertaken by the Social and Health Evaluation Unit of the University of Chester. This focused on the following issues. First the process of collaboration and partnership working between partnership organisation and institutions involved, particularly the project steering group. Second the level of awareness regarding the project amongst the relevant institutions and organisations in the North West of England.

The three pronged focus of the trident yielded a multi method approach which is outlined below.

Outcomes
In terms of outcomes, a questionnaire was designed and distributed to relevant organisations and institutions to establish level of awareness regarding the project and the degree of involvement of key stakeholders. In addition, a partnership assessment was carried out using an established assessment tool and the review of minutes and relevant documentation in order to establish the effectiveness of partnership.

Process
A mapping exercise was carried out in order to provide a detailed account of all the external events organised.

Multi-stakeholder perspectives
Telephones interviews were used to establish the views of the key stakeholders regarding the project and their experiences on the steering group.

Recommendations for practice
The findings led to the number recommendations for future comparable projects regarding the following:

- Partnership aims and working arrangements
- Project management
- Project dissemination
- The Project Steering Group
Evaluation of a Sure Start Citizen Advice Bureau Service

Katrina Stredder and Marie O’Connor
Research Assistants
Social & Health Evaluation Unit
University of Chester

Introduction
Sure Start Local Programmes (SSLPs) have aimed to improve the health and wellbeing of families, with children under the age of 5, in designated areas of deprivation. They are facing, what some consider to be, fundamental changes with the move to Local Authority control. Consequently, the evaluations of local programmes are particularly poignant. The Evaluation Trident was employed in the evaluation of a Citizens Advice Bureau (CAB) service commissioned by a SSLP in the North West of England. The service was delivered via three methods, namely; home visits, ‘drop-in’ sessions and appointments. The CAB Workers offer a wide range of advice and support covering issues such as Benefit entitlement, debt, housing/homelessness, Court Cases and custody cases involving Social Services.

Outcomes
Some difficulties were experienced in identifying the objectives of the service. However, relevant objectives were extracted from the ‘Every Child Matters’ Framework and case studies provided illustration of how these objectives were being met.

Process
The processes involved within the service, including potential pathways for service users and the organisational and functional structure, were identified drawing on existing documentation and stakeholder perspectives.

Multiple Stakeholder Perspectives
Multiple Stakeholder Perspectives solicited included those of service users, service providers, key colleagues, the SSLP Manager and the Chief Executive of the CAB in question. These were obtained via a number of focus groups and individual interviews.

Recommendations for Practice
Data with regard to outcomes was limited but case studies indicated success. This lack of explicit objectives, accompanied by a perceived lack of formality, suggested that a detailed Service Level Agreement was necessary. Nevertheless, the service was deemed successful by all in removing barriers for parents/carers to access services in the first instance and providing proactive means of identifying and resolving issues. This was a result of its non-threatening flexible approach which was recommended to be continued.
Evaluation of ‘Target Word’: a speech and language programme for late talkers

Lamiece Hassan
University of Chester, UK

Introduction
Target Word is a Hanen programme for training parents of children who are ‘late talkers’. Developed by the Hanen Centre in Canada, the program uses a family-centred approach to support linguistic development, the basic premise of which is that parents can become their child’s language facilitator if they learn how. This presentation will describe and report on the findings of an evaluation of the first time that Target Word was run for parents in the UK. In line with the Ellis and Hogard evaluation trident model, this presentation will report on outcomes, processes and multiple stakeholder perspectives.

Outcomes
To determine the impact of Target Word, the evaluation compared the communication skills of parents and their children before and after the course. Pre and post-intervention video clips of parent-child interactions were observed and rated using a novel parent communication rating scale. A standardised communication inventory was used to measure changes in children’s vocabulary scores. Although ongoing data analysis precludes us from citing conclusive findings at this time, initial results appear promising. A fuller account of findings will be presented at the conference.

Processes
The program ran over 10 weeks and comprised eight sessions in total; five two-hour long group sessions and three home visits. Families were recruited to the programme from speech and language therapy department caseloads. Video sessions undertaken on home visits were considered to be a key part of the program, during which parents were filmed interacting with their children and were then given feedback on their interaction style by a speech and language therapist.

Multiple Stakeholder Perspectives
Parents and programme deliverers were interviewed to elicit their perspectives on the program including their perceptions of any outcomes delivered, strengths and weaknesses, and their suggestions for improvement. Transcriptions were subjected to a thematic analysis identifying pre-determined and emergent themes in participants’ discourses. These themes, supported by direct quotations, will be presented at the conference.

Recommendations for Practice
In the spirit of supporting ongoing improvement in forthcoming iterations of the program, recommendations will be produced and presented. It is anticipated that these will relate to both the delivery and content of the program with a UK-based population.
Evaluating a local children's fund programme - lessons learned

Dr Angela Morgan
Senior Researcher
Policy Research Institute
University of Wolverhampton
Email: Angela-Morgan@wlv.ac.uk

This paper provides an overview of the process and key findings from an evaluation of a local Children's Fund programme, conducted by a team of independent evaluators from the Policy Research Institute, University of Wolverhampton. The evaluation was conducted between March 2004 and April 2006, taking an iterative approach in order to acknowledge lessons learned and feed back to the Children's Fund as the evaluation progressed through 6 distinct stages. The paper first provides background information on key policies which provided the impetus for the development of the national Children's Fund initiative, then goes on to detail the local Children's Fund programme in terms of its aims, management structure and projects. The evaluation process is then explained, including the scope, sample and methodologies used. Thereafter, key findings are then presented, providing the basis for subsequent discussion of the chief emergent issues and how they informed recommendations for policy and practice. However, the recommendations are not made uncritically, and it is important to deliberate on how far a relative 'snapshot' evaluation as this can truly impact on the development of children's services, doing this by linking the local evaluation to the national evaluation.

The paper will be of interest to anyone working in the field of children's services, strategic managers, service providers, policy-makers, researchers; indeed anyone with an interest in how evaluation of programmes can inform policy and practice.

This paper relates to the overall conference theme as it discusses the importance of working within a multidisciplinary team of researchers, within the context of statutory, private and voluntary and community sectors, in a policy area fraught with constant change. It subsequently relates strongly to the sub-theme of 'evaluating institutions, programmes and systems' through its not uncritical focus on the evaluation of a local Children's Fund programme, set within a national programme. It will be interesting to compare our approach with those of European positions in the context of services for children.
Evaluating the impact of a UK Government Delivery Body: the Case of the Waste & Resources Action Programme (WRAP)

Dr Barbara Leach
Evaluation Manager
Waste & Resources Action Programme (WRAP)
Email: barbara.leach@wrap.org.uk

Rita Patel
Waste & Resources Action Programme (WRAP)

The Waste & Resources Action Programme (WRAP) was established in 2001 in response to the UK Government’s strategy to manage waste more effectively.¹ A delivery body sponsored by Defra, its mission is to accelerate resource efficiency by creating efficient markets for recycled materials and products, while removing barriers to waste minimisation, re-use and recycling. This symposium paper, which is submitted under stream 5 (evaluating institutions, programmes and systems), will describe the way in which the in-house Policy and Evaluation Team approaches the task of evaluating the impact WRAP has had on resource management.

WRAP undertakes a wide range of activities to achieve its mission including:

- Managing the national TV advertising campaign Recycle Now!
- Issuing capital grants so recycling businesses can improve their infrastructure
- Working with recycling businesses to improve their business plans
- Distributing subsidised home compost bins through our local authority partners
- Working with retailers to reduce packaging

Evaluation activities, which are mainly summative in nature, are diverse and draw on a wide range of research methods including surveys, observational studies, focus groups, secondary data analysis and modelling techniques. These will be described in outline.

The paper will then take a detailed look at how WRAP evaluated the impact of £18 million of funding given to local authorities to run promotional recycling campaigns. The programme carried significant evaluation challenges. Types of monitoring carried out by the funded authorities included observational studies, surveys and waste composition analysis. WRAP then had to aggregate the submitted data and draw conclusions about increases in quantities recycled and the proportion of households participating.

This paper will provide an interesting contrast with fields that have traditionally been covered by the UKES which have tended to focus on social care, health, regeneration, community safety and international development.

Role of participatory approaches to evaluation research in organizational change strategies

José Carlos Suárez
University of Montreal, Canada

Jane Springett
Liverpool John Moores University, UK

Carolyn Kagan
Manchester Metropolitan University, UK

Essentially, participatory approaches to evaluation research (PAER) start out from a recognition that evaluation research develops within a pluralistic society and allows evaluation projects to be built upon the ideas, values and aspirations of those taking part at all levels and throughout the whole evaluation process (Diez et al., 2005). In this sense, Gregory (2000) points out how participation allows people to be aware of the rationality behind the evaluation process, and in doing so, increases their involvement in the implementation of a evaluation research project, both in an effective and efficient way. In other words, PAER reinforce Capacity Building for Evaluation (CBE).

CBE is a context-dependent, organized action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations. Evaluation here is understood as a learning process around a social project evaluated from the perspective of all the stakeholders. It is precisely the real participants in the social project who contribute to understanding and learning about the processes of change underlying the social project and to the development of a new awareness regarding the phenomena under evaluation. Therefore, evaluation becomes an exercise stimulating the appearance of an organizational learning (OL) process, a social process whereby some insight or knowledge, created either by an individual working alone or by a team, becomes accessible to others. The very essence of OL is to adapt to and to transform organizational changes strategies.

Using a precirculated position paper as a starting point we propose to organize a hybrid roundtable and panel/debate to open a discussion to explore the paradigmatic dimensions of PAER in order to establish some conceptual and operational links between CBE principles and organizational change strategies through OL processes.

References


Balancing scientific and political demands: Evaluating and monitoring the implementation of gender mainstreaming in German child and youth associations

Elisabeth Helming
Deutsches Jugendinstitut, Germany
Email: helming@dji.de

German child and youth associations funded by the Federal Child and Youth Plan (KJP) are obliged to implement Gender Mainstreaming (GM). The mandate to account for the gender perspective was explicitly included in the promotion guidelines of the KJP in December 2000. There has been a lot of irritation among the associations by this assignment of the government because of the principle of subsidiarity, though on the other hand there are connecting factors in the endeavours of the organisations to consider the different perspectives of girls and boys in their concepts of youth programmes.

The German Youth Institute has been commissioned by the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth – and this is the objective of the project that is going to be presented – to monitoring and enforcing the implementation of GM in these youth associations, to evaluate the progress of implementation and encourage the organisations in further elaborations of the strategy. This required a method mix: an analysis of the statements on gender mainstreaming made by the services in their annual reports, a standardised questionnaire, and expert interviews as well as group discussions on different workshops with members of several services were instruments of the evaluation.

Monitoring is used by the government to oblige the youth associations to implement GM, while at the same time supporting them. This means quite an ambivalence for the evaluation team. Holding the balance between the requirements of the government and their funding guidelines, the organisations and their claim for autonomy and the demands of a scientific approach and scientific ethics is quite a challenge – evaluation as a critical connection between intervention and activation.
Creating effective evaluation of cultural experiences

Adrienne Pye
Regional Programme Director
Audience Development South East, UK
Email: adrienne.pye@hants.gov.uk

Audience development is the discipline that tries to offer scientific measurement of the impact of art on audiences and participants, and tries to create good communication between the public and the providers.

Responding to some form of monitoring and evaluation has always been an intuitive exercise in the best arts and cultural organisations; it is only latterly that it has become an expressed need and in some cases a funding demand (this very recently, and partly because of pressure from audience development specialists).

It is particularly in participative cultural activity that evaluation methods have had to be created almost from scratch; to evaluate experience is perhaps a step sideways from qualitative research and it is this area that we feel we are still exploring - using methods from visuals to express emotional responses for small children, to video booths in which it is possible to collect unmediated (as far as possible) responses.

Additionally, a great deal of the evaluation work we undertake deals with areas of diversity work, where there are often underlying tensions in organisations that first have to be addressed before it is possible to understand the genuine response to eg educational activity.
Evaluation in Complex Governance Arenas: the potential of Large System Action Research

Professor Danny Burns
Professor of Social and Organisational Learning
SOLAR
University of the West of England, UK
Email: Danny.Burns@uwe.ac.uk

This paper is grounded in the development of systemic action research within SOLAR at the University of the West of England. In it I argue that the creation of dense webs of learning through a process of Large System Action Research offers a solution to evaluators who are grappling with the difficulties of attributing causality in complex governance arenas; the obsolescence of goals, objectives and plans in a fast moving policy environment; the myth of representativeness in diverse contexts, the limitations of snapshots, and the perennial problem that evaluation conclusions don't get acted upon.

The paper explores the way in which emergence, resonance, and playback can underpin a flexible evaluation process that yields valuable systemic insights. A central concern of this work is to ensure not only that it informs change, but also that it creates change, and that learning is generated through that change. The paper draws on three pieces of work to illustrate the systemic concepts that underpin large-scale action inquiry work. One of these, an evaluation of the Welsh Assembly’s Communities First Programme, is explored in more detail. Here I highlight the way in which the emergent nature of both policy and practice is echoed in the emergent design of the evaluation.