Towards an evaluation framework for complex social systems

Diane M McDonald[1,2] **and Nigel Kay**[1]
Information Resources Directorate[1],Computer and Information Sciences[2]
University of Strathclyde

D.McDonald@strath.ac.uk, N.Kay@strath.ac.uk

1. Managing and evaluating complex systems

While there is growing realisation that the world in which we live in is highly complex with multiple interdependencies and irreducibly open to outside influence, how to make these 'systems' more manageable is still a significant outstanding issue. As Bar-Yam (2004) suggests, applying the theoretical principles of Complex Systems may help solve complex problems in this complex world. While Bar-Yam provides examples of forward-thinking organisations which have begun to see the relevance of complex systems principles, for many organisations the language and concepts of complexity science such as self-organisation and unpredictability while they make theoretical sense offer no practical or acceptable method of implementation to those more familiar with definitive facts and classical hierarchical, deterministic approaches to control. Complexity Science explains why designed systems or interventions may not function as anticipated in differing environments, without providing a silver bullet which enables control or engineering of the system to ensure the desired results. One familiar process which might, if implemented with complex systems in mind, provide the basis of an accessible and understandable framework that enables policy makers and practitioners to better design and manage complex socio-technical systems is that of evaluation.

Evaluation approaches, according to Stufflebeam (2001), have been driven by need to show accountability to funders and policy makers examining excellence and value for money in programmes and more recently internally within organisations to ensure quality, competitiveness and equity in service delivery. Another often used broad categorisation is that of requirement analysis, milestone achievement and impact

analysis. Evaluation is therefore a collective term with differing objectives, applications and methods involved. The UK Evaluation Society (www.evaluation.org.uk) defines evaluation as "[a]n in-depth study which takes place at a discrete point in time, and in which recognised research procedures are used in a systematic and analytically defensible fashion to form a judgement on the value of an intervention". Others, however, take a less time dependent approach. For example, on Wikipedia evaluation is defined as "the systematic determination of merit, worth, and significance of something or someone". Evaluation is distinct from assessment which measures and assesses systems be it learning, research or development without making any value judgement. Other, terms often applied to designed systems are validation, which measures 'fitness for purpose' and verification which measures the fitness or correctness of the system itself. Within the context of this paper, we take evaluation to mean the measurement and assessment of the state of a system, in order to form a judgement on the 'value' or 'appropriateness' of the actions being taken by the system. How these different types of evaluation are implemented in practice depends on specific context and objectives. Increasingly a mix and match approach is adopted with appropriate evaluation tools being selected as required. See Snufflebeam (2001) for programme evaluation techniques and LTNI (1998) for general evaluation methods; House (1978) and Stufflebeam & Webster(1980) compare different methods.

Evaluation of Complex Systems is of course not new – educational systems, social initiatives and government interventions are complex social systems where effective evaluation is seen as a key process in measuring success. More recently, there has been an increasing recognition that for evaluation of complex social systems to be effective, the evaluation process must take into account the theoretical understanding of complex systems. For example, Eoyang and Berkas (1998) suggest, "[t]o be effective, however, an evaluation program must match the dynamics of the system to which it is applied." Sproles (2000) emphasises that most systems that are part of a 'test & evaluation' process are actually socio-technical system and he therefore argues qualitative data collection and analysis must play an important role. Practical examples also encompass complexity theory into the design of evaluation programmes (Flynn Research 2003; Webb & Lettice 2005).

Classical evaluation is [a series] of evidence-based value judgements in time. Evaluation determines the merit of a programme or particular artefact or process related to specific stakeholder requirements. It does not necessarily follow that an evaluated system obtains its desired goals or capitalises on its inherent creativity. This leads to the research question - *Can an embedded evaluation system be used as a generalisable framework to improve the success of Complex Social Systems?* To achieve this suggests a non-classical approach to evaluation where the evaluation forms part of a holistic complex feedback and adaption framework. Such an approach would require a number of issues to be addressed. For example, in practice, evaluation within any complex social system tends not to be approached holistically, as individual evaluations are for differing stakeholder; programme level evaluations tend to be distinct from internal evaluation of specific software or processes. While all stakeholder evaluations may show satisfaction, this does not necessarily enable detection or realisation of system potential. Also, the multiple contexts within many complex social systems mean that evaluations in one context do not necessarily transfer to another. And, importantly, evaluation like any measurement changes the system it is evaluating.

This paper reports on a preliminary investigation into the requirements and issues surrounding the applicability of an embedded evaluation framework as a practitioner and policy-maker friendly framework for managing complex social systems. The case study methodology used is outlined in section 2. An analysis of the characteristics of evaluation found to date is provided in section 3. In section 4, a conceptual model for exploring evaluation as a tool for managing complex social systems is present and discussed. This highlights concepts, processes and issues that require further detailed examination if a generalisable evaluation framework for improving the success of complex social system is to be developed. The paper concludes in section 5 by summarising findings and identifying novelty and future research steps.

2. Overview of methodology

To investigate the feasibility of developing an embedded evaluation framework and to develop a conceptual model, a preliminary exploratory study was undertaken. This study examined common themes and differences in evaluation practice across three complex social systems. These case studies were selectively chosen to provide information rich cases that might inform further detailed research. In particular, cases which were using evaluation in a core, unusual or innovative way were chosen. The cases introduced here – criminal evidence systems and learning communities are briefly summarised in section 3.1 below. We also draw on some initial anecdotal evidence from innovation systems, which are also being investigated as part of this work. This part of the study is not sufficiently advanced to merit case study status. The case studies themselves were carried out using a semi-structured interview technique which was supplemented and triangulated using additional documentation relevant to the cases. The interview questions were developed through a literature survey and synthesis and were designed to explore the specification of needs and identification of issues of a new framework approach as well as how evaluation was currently used. Respondents were selected from practitioners and policy makers, as this is the target audience for the research. While the work is ongoing, data from 2 respondents per case study as well as secondary documentation is drawn upon to present preliminary observations.

3. Analysis of the characteristics of evaluation in the case studies

3.1 Overview of the cases studied

3.1.1 Criminal evidence system

In the criminal evidence system, the processes, actors and environment involved from the initial crime scene investigation through the evidence analysis to the presentation of evidence in court proceedings were investigated. The main case used was the Scottish criminal evidence system which is underpinned by Scots Law, which is based on Roman law, combining features of codified civil law with elements of uncodified common law. As the nature of Scots Law – it is an adversarial system - will dictate many of the system processes, the implications of an inquisitorial-based system were also considered. A number of actors are involved in the various stages; police, scene of

crime officers, forensic scientists (defence and prosecution), lawyers (defence and prosecution), Procurator Fiscal, judge, defendant and jury.

3.1.2 Learning communities

Learning communities provided a second case study. These educational systems include learners, tutors or mentors and development staff including educational and technology development. As a focus, the learning communities surrounding the DIDET (www.didet.ac.uk) project, which spanned both UK and US universities and the Transactional Learning Environment project (technologies.law.strath.ac.uk/tle2) were examined. In each of the projects, both the novelty of the learning interventions and the project or transaction based learning scenarios meant that innovative assessment and evaluation techniques were required. Data from more traditional learning scenarios were used to supplement and compare findings.

3.2 Micro and macro evaluation

A variation in the 'level' at which evaluation took place was evident across the case studies. Within the criminal evidence system, evaluation was core to the daily activity of the various professionals involved. For example, police officers normally evaluate finds and other facts at the scenes to direct what is to be collected rather than instigating a blanket collection. Similarly, forensic scientists must weigh up the possibilities that the evidence collected could have been generated through competing scenarios and finally the jury must evaluate the evidence as presented by the various experts. Evaluation is at the core of the criminal evidence system. Similarly, within innovation systems evaluation is part of the day to day processes of creating and developing new products or processes.

In learning communities, traditionally assessment rather than evaluation is much more prominent at the individual learner (micro) level. Evaluation tends to take place at early in and at end of learning programmes to assess the success of the learning intervention and adapt it as required. Increasingly however, an element of learner evaluation is being introduced as part of self-reflection practice within learning communities.

Macro-level evaluation of the system as a whole was also important in both the innovation system and learning communities. Macro-level evaluation of interventions or programmes was however in general independent of the micro-level evaluations, although in qualitative evaluation of learning, there was some use of micro-level reflections of both learners and practitioners.

3.3 The role of context

Two significant issues arose from varying contexts within the case studies. Within the innovation system, a sudden increase in the amount of patents applications – one of the innovation measures – was observed. However, to observers on the ground, the overall innovation system did not appear to have dramatically changed. Further investigation revealed that patenting had dramatically increased because of another separate initiative had spare money which it used to help organisations take out patents. The system as a whole was not intrinsically more innovative as a result although the information available had increased. The problem of differing contexts was also an issue in the

learning community case study. While the supporting technology had been highly successful within the learning context, its design had followed an unorthodox approach. Concern regarding how external peers within the technology development community might judge the developments, initially at least, restricted the ability to evaluate the technological tool.

Thus, as with complex systems in general, the 'measures' can be influenced by different contexts and uses within or outwith the complex system. While evaluation can and perhaps should be ongoing, due to the dynamic nature of complex social systems, any measurement and judgement is only relevant at a specific point in time.

3.4 Complexity

The complexity of the national innovation system gave rise to issues. For example, while evaluation of the various initiatives were always required by the funders and policy makers, a holistic approach was often missing. How different initiatives affected others and how the lessons learnt from current evaluations could be fed back into other ongoing initiatives was often missing. Complexity was also a significant issue within the criminal evidence system with much difficulty in identifying single linear cause and effect relations. For example, it is not uncommon for there to be several competing scenarios as to how particular forensic evidence came to be present. Similarly, in one of the learning communities, lack of uptake of a particular technology was not only due to lack of student eLiteracy training as evaluation first indicated. Subsequent evaluations showed that staff eLiteracy issues were also a strong contributing factor.

So while evidence was found to substantiate the claim that evaluation did have interdependencies, feedback between micro and macro evaluation was in general lacking. Without such feedback, the potential for emergence is extremely limited.

4 Designing an evaluation system for complex social systems

4.1 Conceptual model for evaluation & management of complex social systems

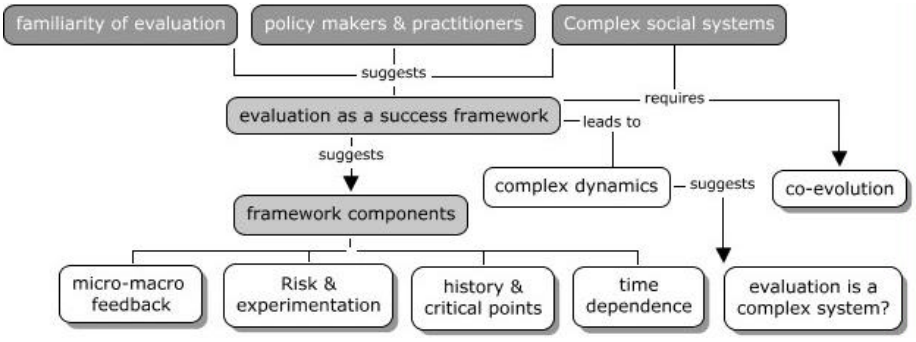

Figure 1: Conceptual model for investigating management of complex social systems using evaluation

From the analysis of the exploratory case studies and synthesis from the literature review, the conceptual model for exploring the design and management of evaluation systems for management of complex social systems illustrated in Figure 1 above was developed. The main features of the conceptual model are discussed below.

4.2 Embedded, co-evolving evaluation framework with feedback

The issues surrounding innovation system evaluation highlighted in section 3.4 illustrate the need for embedding the evaluation framework within the complex system itself; programme end evaluation cannot improve current programmes. Embedding, however, means more than ensuring ongoing evaluation activity. The evaluation framework needs to be able to detect changes and new system potential. The evaluation must also remain relevant and timely; as the system changes through experimentation, the evaluation must also adapt, as the evaluation measures may no longer reasonably represent the desired characteristic, as the patent example of section 3.3 illustrated.

For evaluation to be an effective management tool, the results should be fed back to inform general system behaviour and specific interventions. While this means that evaluation will change the system, the converse should also hold – system changes require change in the evaluation criteria and practices. The evaluation system should co-evolve with its complex social system.

Applying complex systems thinking to evaluation also suggests the potential advantages of linking micro and macro level evaluation feedback. While such feedback helps build buy-in and a sense that individuals can make a difference, from a complex systems perspective, it is such micro-macro linkages that leads to novelty – the emergence of sustainable new patterns of activity (McDonald & Weir 2006). The potential of micro macro feedback requires further exploration.

4.3 Evaluation history and critical evaluation point identification

The case studies highlighted the run up to new funding opportunities or new development phases as critical evaluation points from the practitioner and policy-maker perspective. Such points tend to be snapshots in time and do not necessarily identify developing patterns. For a fuller picture, evaluation requires to be ongoing. Continuous evaluation however is both a large overhead and potentially counter productive. For example, one of the case studies reported evaluation skewing the development - the evaluation schedule drove development rather than the actual system requirements. Additionally, participating in evaluation was viewed as time consuming. The richest feedback was obtained when participants understood how the information being gathered would be fed back to improve the system.

While impact type evaluations aim to provide reassurance of quality and meeting of objectives, it is where evaluation outcomes differ from expectations that will be indicative of the future direction of the complex system. It is at these points that either corrective intervention is required in the case of undesired behaviour or new potentials may be capitalised upon. Identification of these bifurcations points however requires knowledge of existing patterns. The sharing – often informally - of reflective practice of participants own experiences were viewed as an effective way to identify emerging phenomena and patterns. Such a second layer communal reflection may help minimise personal biases, enabling patterns unseen by one participant to be identified by another

due to differing perspectives. The use of historical evaluation records for pattern identification and scenario mapping also requires further investigation.
Thus, while change processes often dictate evaluation points, a second layer of reflective evaluation potentially offers a more complete picture, enabling critical points to be identified and interventions designed. A framework which supports the feedback of these evaluations in a timely manner is critical. Further work is required to help identify the parameters associated with optimal evaluation and feedback points.

4.4 Risk and experimentation

Complex adaptive systems by nature 'experiment', with various adaptions being selected over time. While such systems can be highly creative, this is not always desirable. Successful management of systems often involves a trade-off between experimentation and the risks of undesirable behaviour. Within the innovation system, a certain amount of risk was viewed as highly desirable, but this needed to be balanced against overall productivity. Within the criminal justice system, with its need to provide detailed forensic evidence, the scope for experimentation was extremely limited. Experimentation, such as new forensic techniques, was usually thoroughly tested outside criminal proceedings before they were accepted. Adaption and variation does take place however, with new combinations of techniques being used, which must 'compete' against the other sides methods. This suggests the possibility of developing an evaluation framework based on different classes of risk-experimentation trade-off.
Classifications of a system may also vary depending on conditions. For example, in the criminal investigation case study, it was reported that the normal 'rules of the game' may change in extremely distressing, high profile cases such as child murders. Suddenly, the realisation that the evidence will be reported to the child's parents greatly increases the need to minimise risk. This suggests a dynamic trade-off between risk, and experimentation. The potential of a classification, based risk-experimentation trade-off, which takes into account system purpose, constraints, contexts and dynamics requires further investigation.

4.5 Discussion

As Wang et al (2004) observe (in relation to learning environments) the complex nature of artefacts in a "threat to classically ideal evaluation design where all the variables are well controlled except the [facility under evaluation]". But these interactions that we have identified here hint at another important point worth consideration – is an evaluation system designed with complexity in mind in fact a complex system itself? If so, this leads to the possibility that self-organisation and emergence could occur. Again, self-organisation and emergence may not potentially follow the desired path although the effect of coupling through embedding with the social system itself may provide the guidance required. This is analogous to Ottino's (2004) suggestion of "intelligently guid[ing] systems that design themselves in an intelligent manner". Similarly, unpredictability which we are trying to solve may potentially occur, but coupling may keep it pointed in the right direction. Detailed research on the true dynamics of complex evaluation systems is required.

5 Summary, novelty and next steps

In this paper, we have presented insights from a preliminary investigation of evaluation in Complex Social Systems which examined correlations and differences between three differing complex social systems. The aim of this study was to develop a conceptual model of evaluation and control of complex socio-technical systems which could then be used for further detailed study ultimately to produce a policy and practitioner relevant evaluation and control framework for complex social systems. The preliminary insights derived were: (i) Embedded evaluation system is required which co-evolves with its complex social system, coupling micro and macro level evaluations, (ii) Evaluation may be most effective when there is a feedback between micro and macro evaluation; (iii) Context varies both within and outwith the evaluation system perturbing the evaluation and (iv) Different classes of evaluation system may be appropriate to deal with trade-off between purpose, constraints and experimentation.
The novelty of this work lies in the setting of a future research agenda for exploring evaluation as a success mechanism for complex systems and the identification of the additional issues which arise when evaluation is applied within a complex systems framework. The next steps are to undertake a detailed investigation based on conceptual model developed (Figure 1), identify critical evaluation points and explore the complex dynamics of evaluation. This will ultimately lead to a policy maker and practitioner friendly complex evaluation and management framework for complex social systems.

References

Bar-Yam, Y., 2004, Making things work: solving complex problems in a complex world, NESCSI , Knowledge Press.

Eoyang, G.H. and Berkas, T.H., 1998, Evaluating performance in a CAS.

Flynn Research, 2003, Complexity Science: A conceptual framework for making connections – Denver. Flynn Research.

House, E.R., 1978, Assumptions underlying evaluation models. Educational Researcher Vol 7, num 3, pp4-12.

LTDI, 1998, Evaluation Cookbook. Learning Technology Dissemination Initiative - http://www.icbl.hw.ac.uk/ltdi/cookbook/cookbook.pdf (accessed 14/08/2006)

McDonald, D.M. and Weir, G.R.S, 2006, *Developing a conceptual model for exploring emergence – ICCS 2006* (Boston).

Ottino, J.M., 2004, Engineering complex systems. Nature Vol 427, p389.

Sproles, N., 2000, Complex Systems, Soft Science, and Test & evaluation Or 'Real men don't collect soft data' - Proceedings of the SETE2000 Conference.

Stufflebeam, D.L. 2001, Evaluation Models New directions for evaluation, num89.

Stufflebeam, D.L. and Webster, W.J., 1980, An analysis of alternative approaches to evaluation. Education evaluation and policy analysis, Vol 2, num 3, pp5-19.

Wang, H.-C., Li, T.-Y. and Chang, C.-Y., 2004, Analyzing Empirical Evaluation of Advanced Learning Environments: Complex Systems and Confounding Factors. IEEE Learning Technology Newsletter, Vol 6, num 4, pp39-41.

Webb, C. and Lettice, F., 2005, Performance measurement, intangibles, and six complexity science principles - International Conference of Manufacturing Research, Cranfield.