Evaluation in Software Engineering: ROI, but more than ROI

Barbara Farbey **Anthony Finkelstein**
Dept. of Computer Science
University College London
Gower Street,
London WC1 E 6BT, UK
44 (0)20 7387 7293
b.farbey@cs.ucl.ac.uk | a.finkelstein@cs.ucl.ac.uk

ABSTRACT
This paper presents a short survey and analysis of a variety of evaluation methods, which are potentially useful for strategic software design. The methods described address a broad range of concepts of value. The paper complements previous economics driven software engineering research work on standard accounting techniques. A summary account of the underlying principles of each method is given, the methods are classified according to these principles and a way of choosing an appropriate method described.

Keywords
Strategic Software Design, Evaluation, Participative evaluation

1 INTRODUCTION

There are many techniques, models and philosophies which can be adduced in the evaluation of software engineering and design. Previous work in the field has discussed standard accounting techniques, drawing attention to design and engineering economic activity, to be judged in the end by its contribution to business value. ([2],[6],[7],[12],[18]). This paper complements that work, presenting short, summary accounts of a range of evaluation methods for use in situations where market values do not necessarily dominate.

Many of the methods were developed in the context of social programs, for example health or education programs. They are, therefore, designed to address the values and needs of a multiplicity of stakeholders, rather than the narrower focus of shareholder value, which underpins for example, the justification for Net Present Value calculations.

They are of interest to software engineers who more often than not are faced with having to elicit, explore and then reconcile differing sets of values, at least sufficiently for the practical purposes of software design.

Section 2 sets out a selection of the principal methods. They have been chosen as exemplars of their kind, where possible on the basis of accepted use. Unfortunately this latter is not always possible. There are if anything far too many methods, roughly one per consultant. Few have been tested on any scale in industrial conditions. In each case a primary reference is given where details of the method can be found.

Section 3 describes a classification of evaluation methods, taken from the discipline of Evaluation and due in the first instance to House [13]. Using the classification it is possible to understand the roots of the different approaches and therefore begin to choose a method which will suit the purpose at hand, especially in situations where economic value is not the primary consideration. Section 4 shows how, within the broad classes, the choice can be narrowed following a method devised by Farbey et al [10].

2 METHODS

A wide-ranging discussion of appraisal and evaluation methods can be found in [9]. Here we give thumbnail sketches, with references to further more detailed, or seminal, accounts.

Quantitative and comparative methods

Cost-benefit analysis and hedonic models
Cost-benefit analysis (CBA) is a variant of Return on Investment (ROI) in which intangibles, both costs and benefits, are assigned monetary values and form part of the ROI calculation [8]. A CBA may, for instance, attempt to attribute a monetary value to human life for medical budgeting or insurance purposes. Cost benefit analyses are widely used as input into public decision-making processes and increasingly in investment appraisal for IT investment. A detailed discussion is provided in [9].

Hedonic pricing is a methodology for valuing non-market goods that affect market prices. It is used, for example, in estimating variations in house prices to reflect the value of environmental features, or more generally to estimate the environmental impact of economic activity. The results can then be fed into a CBA. For a brief introduction see [4].

Cost-benefit analyses are widely used, but there are reservations as to their value where there is not widespread agreement as to the basis of the quantification procedures.

Information Economics
Information Economics [17] extends Cost Benefit Analysis by incorporating three extra calculations: *value linking*, *value acceleration* and *job enrichment*.

Value linking extends the basic CBA/ ROI process to look for the costs and benefits of organizational changes which follow the introduction of a new system, but which are not the immediate targets. Value acceleration brings into account the future effects of an investment. For example the introduction of an extranet may allow a knowledge management system to be introduced earlier than it might otherwise have been. Job enrichment includes as part of the benefit calculations, individual and organizational learning and increased skills.

Information Economics is complex and expensive, and therefore cumbersome for small projects. Against that it is very much in tune with the contemporary emphasis on learning and knowledge management.

Qualitative and exploratory methods

Value Analysis
Value Analysis [19] is an exploratory technique, which tries to assess the incremental value of the outputs of a proposed system, principally its value to decision-making and decision-makers. A number of procedures are involved.

First an agreed estimate of the value of a proposed system is established, via for example, a Delphi procedure. Next, a working model of how the system will work in practice is constructed, for example a prototype, or possibly a role-play exercise which assumes the system outputs. Potential users and managers use the model to help improve the estimates of value and to decide whether further improvements are required. If so, there is a further decision as to whether the benefit of including them would outweigh the cost. The process is iterated until a system evolves which is regarded as satisfactory for the cost.

The advantages of Value Analysis are that the process establishes agreed values for intangible outputs, and that decision-makers are involved throughout, which tends to build confidence in the eventual result. Moreover, value can be expressed either as utility, or in money terms. In the latter case the results can be fed into a standard ROI calculation. The disadvantages are that it takes time and money and the accuracy of the intermediate stages is not always sure.

Value Analysis appears to be an attractive way of evaluating requirements of software systems and, because of its incremental nature, new features in product families.

Multi-objective, multi-criteria methods
This family of methods explicitly recognizes the existence of many points of view and more than one set of values in the decision to invest in a system. Multiple-objective, multi-criteria (MOMC) methods do not rely on monetary measurements of value, instead they work via an iterative procedure to establish preferences and utilities [15]. The precise procedures, for example the weighting of individual preferences or the composition of evaluating group, differ from one variant to another. Nevertheless the general thrust is to *"find a function from a set of individual preferences to a social preference order"* [5], i.e. MOMC methods have their roots in Social Choice Theory and Multi-attribute Utility Theory.

As with Value Analysis, a prime advantage of MOMC methods is the opportunity they provide for exploration both of the problem in hand and the views and preferences of those affected

Fourth Generation Evaluation
Fourth Generation Evaluation (4-G) is in part a reaction to earlier, quantitative and so-called "positivist" methods [11]. 4-G methods proceed from an entirely different set of premises, being interpretivist in approach and depending on a view of reality as social construction. The evaluator, or researcher, is no longer an objective or expert outsider, but part of the scene. The evaluator functions as an organizer and, if required, as coach. The evaluation does not yield a unique "truth"; instead many "truths" are juxtaposed and explored until the problem is better understood and/ or a basis for action emerges.

Although Guba and Lincoln's methodology is the most extreme of the interpretivist methodologies, and difficult to envisage as acceptable in a field of engineering, a more moderate, interpretivist, situated evaluation has been the subject of growing interest in IS evaluation [3],[20]. The shift of the locus of control from an heroic evaluator to the participants in a programme, is significant, particularly in strategic software design where the parameters are fuzzy and the properties of the system strongly emergent. Coming to understand the values of other participants and giving them due voice is often the most valuable part of an evaluation procedure. That said, the methods are not suitable where alternative solutions, projects and investments have to be directly compared.

Socio-technical evaluation
Software Engineering creates socio-technical systems: systems of people and machines. Socio-technical design has, with mixed fortunes, been used in industry for almost half a century. It's primary focus is "*on the design of work*

systems to improve the welfare of employees" and its philosophical basis the "Humanistic Welfare" paradigm. Land [16] has recently suggested incorporating the principles of socio-technical design into the evaluation of information systems. He recognizes that in order for any evaluation to work it must be acceptable to those who are responsible, and that for the most part business management requires "the numbers". He therefore suggests as a first step the incorporation of an employee perspective into the commonly used "Balanced Scorecard" of Kaplan and Norton [14]. The Balanced Scorecard, itself a departure from the overwhelming emphasis on shareholder value in the 1980's, considers four perspectives:

1. The financial perspective – how do we look to our shareholders?
2. The customer perspective – how do we look to our customers?
3. The internal perspective: how efficient and effective are we?
4. The learning and growth perspective – how innovative are we?

Following Land, there would be a fifth perspective:

5. The employee perspective – are we improving the quality of working life?

Like MOMC, an evaluation which takes a traditional socio-technical stand, would pay attention to many points of view. It would also raise as issues:

- the process of evaluation which should itself be socio-technical
- the locus of control
- the criteria for evaluation
- the assumed value consensus.

Although it does not prejudge these issues, it is naturally a collaborative and situated process and the locus of control would most likely shift to the work-group, rather than management or an evaluator.

Summary

There are, then, a variety of methods and approaches to evaluation. They are not only different in detail, but in the fundamental assumptions they make about the world and in the purposes they serve. Choosing a method is therefore partly a question of understanding and choosing a set of assumptions and partly a question of choosing a tool that is acceptable and will do the job intended. In the next section we examine a classification of methods and approaches which helps to illuminate the assumptions. Following that we present a way of choosing a method, based on the degree to which business aims and technological capabilities are understood.

3 EVALUATION TYPES AND MODELS

The first step in choosing a method of evaluation is to understand the underlying model. House [13] has devised a taxonomy of evaluation models which develops the major classes of evaluation approaches as they were in 1980. House's taxonomy is reproduced in Table 3, modified to allow for the application to Software Engineering.

The type or model for evaluation appears in Column 1. It should be remembered that the vocabulary is taken from the discipline of Evaluation and is therefore technical to that discipline. For example, the first type in the table is "Systems Analysis" to which we have added Accounting.

In this category we mean to include models of evaluation which rest on a series of agreed accounting conventions, assumptions that there are causal effects which are known or can be discovered by measurement, quantitative analysis and that well known optimization techniques will hold good. It is, as Butler et al remark, the situation where the data is "good"[1]. Historically, these models maintained a system's view of the world, hence the term "systems analysis".

"Goal free" evaluation is free of goals in the sense that the evaluator is not making the final decision. For example, consumer magazines often publish comparative studies of consumer goods in which the features of, say, washing machines, are listed and compared. Apart perhaps from a non-binding "best buy" recommendation, it is the consumer, not the evaluator who has the final say.

"Art criticism" has as its essence the employment of an expert who makes a judgement on behalf of a client, or readership, on the basis of expert knowledge. In software terms, it is not uncommon for an expert, or consultant, to be asked to provide a critique of a new system, especially where the organization does not itself have the in-house expertise.

"Quasi-legal" evaluation is often mirrored in systems development by the project champion acting to persuade people of the virtues of the system, and an "anti-champion" arguing against.

The remaining types are familiar in the context of systems development and are largely explained by the remaining description in the later columns.

Although a plethora of methods has appeared in the discipline of Evaluation since the publication of the table, the typology needs very little extension. Most "new" methods fit quite neatly into one or other of the House categories. However to take account of recent developments, we suggest the addition of two further models and one extra distinguishing feature.

The first extra category comprises collaborative and participative models. These overlap to a large extent and in particular they have in common that the balance of responsibility for the evaluation is shifted from the "evaluator", considered as an external, objective or expert observer, to project participants. As we have said above, the role of an outside evaluator, if there is one, is more like that of a facilitator or coach. The second category is socio-

technical evaluation, as put forward by Land [16] and which is currently being developed by Farbey, Land and Targett.

By examining the table it is possible to choose a broad approach to any particular development. Thus if the primary purpose of an evaluation was, say, to choose one of a number of COTS offerings, it would be sensible to do a goal free evaluation to establish the principal features of interest, and weed out some of the unlikely contenders. This could be followed by a review of the remaining options by the professionals concerned. Or if the proposed system were for a single, unique group, employing a collaborative evaluation approach would allow the voice of the customer/ user to be clearly heard.

4 MATCHING THE EVALUATION METHOD TO A SITUATION

Evaluation is done for many purposes and in many different contexts [10] presents a procedure which suggests what methods or approaches to evaluation are suitable, for a given organizational purpose and context.

The procedure has three stages:

First the organizational factors which affect the choice of a method are identified. Five groups of features are identified in [10]:

- the role of the evaluation
- the decision environment in which evaluation takes place
- the system characteristics
- the organizational characteristics
- the specificity with which cause and effect between an investment and its benefits are linked.

The information is used to locate each feature on a 2x2 matrix. The matrix is shown in Table 1.

The columns represent the degree to which the effect of the system is predictable. A new technology, or the introduction of technology into a new business area, for example e-commerce, are radical innovations and their impact highly uncertain.

The rows represent uncertainty in the business area. The business context may be turbulent, as with dot coms., or it may be that the requirements are not known in detail. In practice this is usually reflected in the degree to which the evaluation procedure is constrained, or *ad hoc*.

		Role of Proposed System	
		Conservative	Radical
Evaluation constraints	Well defined		Security portfolio [2]
	Fuzzy	Strategic flexibility	Collaborative commerce, product lines

Table 1. Characteristics of the situation, with examples.

Second, the information, such as that provided by House's table, is used to characterize the methods in the same way and the methods entered onto a similar matrix. The classification of the methods described in Section II used in [10] is shown in Table 2.

Third, the two matrices are overlaid and the method chosen according to the match produced.

Although this is still only a rough guide to the final choice of a method, it is a systematic way of tackling the problem, narrows the choice and promotes awareness of the options available to mangers.

As a project moves through the lifecycle, it is possible that several methods will be required. For example in the conceptual phases of a project, where the business goals are unclear, MOMC could be called into play, followed by an ROI when a consensus emerges. Or an initial ROI calculation for a well-understood product could be extended using Value Analysis for similar products in a product line. The broad rule is to push "up" and "left" across the matrix, using the methods, as appropriate, to decrease uncertainty in the business area (up) or in the knowledge of how the technology will operate (left).

		Role of IT	
		Conservative	Radical
Evaluation constraints	Well defined	ROI	CBA Experimental methods
	Fuzzy	MOMC	Information Economics Value Analysis

Table 2. Classification of methods [10]

5 SUMMARY

[1] has correctly identified the problems which arise when a rigid model of procedure is stamped onto a problem in evaluation which it doesn't quite fit. This paper has sought to present in abbreviated form further options for evaluation and a guide to choosing between them.

The principal recommendation is that although in a business context the economic implications of strategic software design are ultimately paramount, other approaches can and should be used, where appropriate in conjunction with accounting methods, to add significant value to the final outcome. The purpose is not so much summative, as envisioned by accounting procedures. It is to draw out and find agreement on the worth of the proposed system and the procedures by which it is produced.

REFERENCES

1. Butler, S., Jha, S. and Shaw M. (2000) *When good models meet bad data: applying quantitative economic models to qualitative engineering judgements*, Position Paper for EDSER-2, On-line at http://www.cs.virginia.edu/~sullivan/edser2/index.html
2. Butler, S., Chalasani P., Jha, S., Raz O and Shaw, M.

The potential of portfolio analysis in guiding software decisions, School of Computer Science, (1999) Carnegie Mellon University, and Department of Computer Science, Arizona State University (Chalasani)
3. Diez M.A. and Esteban, M/S. *The evaluation of regional innovation and cluster policies: looking or new approaches*, On-line at http://www.europeanevaluation.org/pdf/2-2_diez.pdf
4. Ecosystem Evaluation, *Hedonic Pricing Method in Ecosystem Valuation*, Methods, Section 3, On-line at http://www.ecosystemvaluation.org/hedonic_pricing.htm
5. Elster J. and Hylland A. *Foundations of Social Choice Theory*, (Cambridge University Press, Cambridge, 1986)
6. Erdogmus H (1999) Comparative evaluation of software development strategies based on Net present Value, EDSER-1, On line at http://www.cs.virginia.edu/~sullivan/edser1/index.htm
7. Erdogmus H. (1999) Valuation of Complex Options in Software Development, EDSER-1, On-line at http://www.cs.virginia.edu/~sullivan/edser1/index.htm
8. Emery J. Cost Benefit Analysis of Information Systems, SMIS Workshop, no.1(1973)
9. Farbey B., Land F.F. and Targett D. *How to Evaluate Your I/T Investment: A study of methods and practice*, (Oxford, 1993) Butterworth Heinemann
10. Farbey B., Land F.F. and Targett D. Evaluating Investments in IT, *Journal of Information Technology*, **7**, (1992) 109-122
11. Guba, E.G. and Lincoln, Y.S. *Fourth Generation Evaluation*, (Thousand Oaks, 1989) Sage
12. Harrison W., Raffo D., Settle J., and Eickelmann N. Adapting Financial Measures: Making a Business Case for Software Process Improvement, *Software Quality Journal* ,**8**, 3 (1999) 211-231
13. House, E.R. *Evaluating with validity*, (Beverly Hills 1980) , Sage Publications
14. Kaplan R.S. and Norton D.P. *The Balanced Scorecard: translating strategy into action*, (Boston, Mass. 1996) Harvard Business School Press
15. Kenny, R.L. and Raiffa, H. *Decisions with multiple objectives: preferences and value tradeoffs*, (New York, 1976) John Wiley and Sons,
16. Land F.F. Evaluation in a socio-technical context, in *Organizational and Social Perspectives on Information Technology*, (eds.) Richard Baskerville, Jan Stage, Janice I. DeGross, IFIP TC8 WG8.2 International Working Conference on the Social and Organizational Perspective on Research and Practice in Information Technology, (Aalborg, Denmark, June 9-11, 2000) Chapter 8
17. Parker, M.M., R.J. Benson and H.E. Trainor, *Information Economics: Linking business performance to information technology*, Englewood Cliffs., 1998) Prentice Hall
18. Raffo D, Settle J and Harrison W (1999) *Estimating The Financial Benefit And Risk Associated With Process Changes*, EDSER-1, on line at http://www.cs.virginia.edu/~sullivan/edser1/index.htm
19. Rivard E. and Kaiser K. The benefits of quality IS, *Datamation*, (January, 1989), 53-58
20. Smithson, S. and Hirschheim, R. (1998) Analyzing information systems evaluation: another look at an old problem, *European Journal of Information Systems* **7,** 3, (1998) 158 - 174

ACKNOWLEDGEMENTS
This paper extends to Software Engineering earlier work in Evaluation by Barbara Farbey with Frank Land and David Targett.

Model	*Major audiences or reference groups*	*Assumes consensus on*	*Methodology*	*Primary desired outcome*	*Typical Questions*
Accounting/ Systems Analysis (note: "systems analysis" was the name given to a range of approaches originating with the RAND Corporation)	Management, regulatory bodies, government	Goals, known cause and effect, quantification procedures and quantified variables	- Return on Investment Analyses, including Net Present Value, Options Analysis, Cost-Benefit Analysis - Quantitative modeling	Efficiency	- Ex ante: What return on Investment is expected? - Ex post: Have the expected effects been achieved? - Both: Can the desired effects be achieved more economically? - Which are the most efficient investments?
Decision making Operational Research	Decision makers, especially project management, administrators	General goals, criteria	- Surveys, questionnaires, interviews - Quantitative modeling including simulation, math. programming, experimentation	Effectiveness, quality control	- Ex ante: Will the system be effective in use? - Ex post: Is it effective? - Ex post: Which parts are effective?
Behavioural objectives	Managers, psychologists, users	Prespecified objectives, quantified outcome variables	- (establish) behavioural objectives - achievement tests	Productivity, accountability	- Is the system achieving the objectives? - Is the system producing? - Are people behaving appropriately to the system?
Goal free	Users , Consumers	Consequences, criteria	- Requires bias control, logical analysis, attention to *modus operandi*	Consumer choice, social utility	- What are *all* the effects?
Art criticism	Connoisseurs, consumers	Critics, standards	- Critical review e.g. by consultants	Improved standards, heightened awareness	- Would a critic (skilled outsider) approve this system? - Is stakeholder appreciation of the issues increased?
Professional review	Professionals, public	Criteria, panel procedures, panel composition	- Review by panel, self-study	Professional acceptance	- How would professionals rate this system?
Quasi-legal	Jury	Procedures, judges	- Quasi-legal procedures	Resolution	- What are the arguments for and against this system?
Case study	Client, practitioners	Negotiations, activities	- Case studies, interviews, observations	Understanding of diversity	- What does the program look like to different people?
Collaborative evaluation (added to House's original)	Practitioners/ users	Who is to be heard	- As chosen by the participants	Improved awareness	- Who is "we"? - What are we trying to do? - Where are we now? - What do we have to do in the future to achieve the objectives?
Socio-technical evaluation (added to House's original)	Employees with management	System boundary, desired behaviour, legitimacy of process, criteria, place of work in human growth	- Socio-technical. - Participative and collective decision making: continuing, iterative	Self-designed autonomous work group; shift in locus of control	- Are we contributing to improving the quality of working life?

Table 3 A Taxonomy of evaluation approaches. After: ER House, Evaluating with validity [13]