

OPTIMISING 'BANG FOR BUCK' IN EVALUATION

Evaluation in large organisations can be costly and complex as they involve human resources (time, training, energy) money (travel, accommodation, software, hardware, communications) and often reprioritising of other organisational functions. Wholey (2004) confirms that evaluation can be costly and yet can yield little useful information, unless steps are taken to test design and utility of data beforehand. Costly evaluation functions that deliver little relative value are obviously counter productive for the organisation, and yet are common.

Stufflebeam (1999) developed the Meta-evaluation Model for evaluating the evaluation. The strength of this model is its thoroughness in the identification of the variables that ensure the integrity, reliability and validity of the evaluation process, as well as the quantifying the evaluation through a set of simple metrics. Experience dictates that transposition of Stufflebeam's Meta-evaluation Model to smaller evaluations, is problematic because such evaluations cannot be adequately measured against all the standards and components. This paper will extend on the Meta-evaluation Model by assigning an alternate set of metrics for analysis of evaluations to give a holistic yet simpler perspective on their value. The basic premise is that unless the evaluation can be utilised effectively, then it has limited impact ('bang for buck'). Rossi et al (2004:411) contend that the worth of the evaluation 'must be judged by its utility'. Rossi et al (2004) further generalises that there is much disappointment about the utilisation of evaluations, which is attributed to their limited direct or instrumental use. Evaluation utility, as the underlying theme of this paper, will be expressed as a product of the motive for the evaluation, the evaluation design, the potential impact of the evaluation and validity of recommendations.

Motive for Evaluations

The motive for the evaluation is the overall purpose as to why the evaluation is undertaken. Chelimisky (2006) highlights that evaluation purposes are not mutually exclusive, nor should they be treated in isolation. Evaluation utility has undergone continuous change with theorists proposing different approaches. The assumption is made that the later applications build upon previous with each generative approach is more refined. Based on evolving thought and practice, a proposed hierarchy of motivation for evaluation with representational values is listed below:

- 1 Compliance
- 2 Performance against program aims
- 3 Quality of the program
- 4 Process improvement
- 5 Cost-benefit of the program (eg ROI models)
- 6 Organisational decisions (program continuation decision)
- 7 Knowledge generation
- 8 Organisation learns and grows

The initial level is a measure against compliance. Compliance is against policy, legislation or other organisational governance requirements. Compliance is considered at the lower end of motive, as it is the least likely purpose that could value add to the organisation. Henry (2000) argues that the majority of evaluations at this level tend to focus on minor issues of irrelevance.

The second level, is a measurement against performance of aims. For Rist (1994), the majority of evaluation data utilisation is geared towards an accountability measurement. Aims directed evaluation can determine for the program manager whether the program is on track or has met its stated aims

The third level is a measure of the quality of the program. Quality is a commonly stated evaluation motive within the education and training environment, however many evaluators have difficulty contextualising it. Stake and Schwandt (2006:404) define quality as the 'unique things' within quality by identifying them as 'notions of merit, worth and significance'. The challenge for the evaluator measuring quality is to be able to reference some form of scale and identify the merit, worth and significance within the evaluation process.

The fourth level is concerned with the improvement of the program. Nevo (2006) confirms that the improvement focus is tied in with the ongoing requirements for organisations to be both innovative and undertake modernisation. This level of application is typical for personnel interested in improving the effectiveness of the current and future programs.

The fifth level is the cost-benefit analysis of the program. Schmidt (1998) advocates the cost-benefit analysis evaluation, as organisations have limited resources to support their training endeavours and need to quantitatively justify them to management. Schmidt (1998) strongly recommends that evaluators learn financial techniques to enable them to conduct this level of evaluation.

The sixth level relates to helping make significant or strategic decisions within the organisation. Schadlock and Thornton (1988) highlight that the main outcome of an effective evaluation is if relevant decisions are made within the organisations which lead to action to be taken. The evaluation, according to them, has the potential to reduce uncertainty, application and reasonableness. The evaluator must be circumspect of the importance of evaluation data in the decision making process, because as Bemelmans et al. (1994) highlight, it is only one of many contending factors that lead to the decision.

The seventh level is where the knowledge from the evaluation can transcend the organisational boundary into a wider community of practice. According to Rossi et al (2004) some evaluations that describe the nature and effects of the program, can contribute to the overall knowledge generation of the program. The audience for this knowledge are not only those immediately associated with the program but could be disseminated through broader audiences of scholars and policymakers.

The final level is organisational learning. King (2004) is convinced that the evaluation should have a strategic focus through increasing the organisation's ability to learn from the evaluation data, and to grow through building their capacity from the data. Both this learning and capacity building focus according to King (2004), could be achieved through active participatory evaluation methods.

Potential Impact of Evaluations

The impact of the evaluation is the causal effect of the evaluation on the organisation. The potential impact can be determined from a judgement of how the recommendations can affect the organisation. Ideally, if the design and motive for utility is effective, then there should be a close correlation with the impact of the evaluation. Organisational structures, constraints and influences within the evaluation processes can affect the impact of the evaluation. Rogers and Williams (2006) caution that evaluation is heavily subjected to influences by the organisational dynamics. If the dynamics within the organisation that prohibit organisational change can be overcome, then, evaluation can be an effective tool in facilitating the improvement process.

A proposed hierarchy of potential impact for evaluation with representational values, aligned to the motive, is listed below:

- 1 No apparent impact (primarily or totally used as compliance)
- 2 Comparatively minimal impact (measurement against program aims)
- 3 May lead to impact (through enhancing quality systems)
- 4 Important to the program (improvement of the process)
- 5 High level of importance to the program (aligned to organisational decision making models)
- 6 Critical to the program/organisation (major decisions about the program or organisation).
- 7 Knowledge generation (knowledge from evaluation applied across programs and contexts)
- 8 Organisation can learn and grow from the evaluation findings.

Evaluation Design

There are numerous evaluation designs to apply into the evaluation process from the traditional quantitative approaches through to modern qualitative collaborative methods. Walker and Wiseman (2006) highlight

that good design results from careful matching of the aims and objectives against the methodology. They also caution that the design of the evaluation is often prespecified within the original intent of the commissioning or stakeholder authority.

To measure the impact of design as a function of the value for evaluation, this study will apply the design used by the Royal Australian Air Force to evaluate its' training courses. The basic premise of applying this design is if the elements of the preferred organisational evaluation design are adhered to, then the evaluation is likely to be effective. Untested within this premise is the validity, reliability and robustness of this design.

The Air Force evaluation design follows a four step process, reflecting common evaluation design elements. The steps are designed as a lock-step evaluation process, with an emphasis on simplicity for personnel without formal evaluation education and experience to implement.

Step 1 requires the evaluator to develop an evaluation plan. The development of evaluation aim and questions should be negotiated with the stakeholders. The plan also defines the scope of the evaluation; the appropriate tools and methods; a measurement strategy how many and who to survey and to determine what to do with the results. Step 2 requires the evaluator to collect and collate information. Step 3 requires the evaluator to analyse evaluation information/data.

Step 4 is the reporting of evaluation and monitoring of outcomes with commissioner/stakeholders.

The steps with ascribed values of a simple numeric for each of the elements is represented in the table below. The table assumes equal values for each design element with the summation of the elements representing the value of the design metric.

Step 1					Step 2		Step 3	Step 4	
Eval Aim	Eval Q	Data sources	How and when data	MGM T of results	Collect data	Collate Data	Analysis	Reporting	Monitor
1	1	1	1	1	1	1	1	1	1

Table 1 Summary of the Organisational Design steps and components and assigned values.

Percentage of recommendations implemented/endorsed

Percentage of recommendations implemented is a metric applied in recognition of the importance of the relevance of the evaluation findings. Hendricks and Papagiannis (1982:121) reinforce that 'recommendations are one of the most critical products of an evaluation'. If the recommendations align to the social, political, structural and resource framework of the organisation, then a reasonable assumption is that a high percentage of recommendations should be implemented or at least endorsed by the stakeholder/commissioner. A low implementation rate or endorsement by the stakeholder/commissioner would be indicative of flaws within the evaluation planning, design and/or the framing of the evaluation findings. The recommendations are expressed as a decimal fraction with this metric.

Formula for Determining 'Bang for Buck'

In determining a formula for measuring evaluation it must be acknowledged that figures are not precise because of the numerous and complex variables they represent. The application of metrics, however can give comparative values to analyse organisational phenomena. Phillips (1998:214) reinforces the value of metrics by stating: 'they tend to be as accurate as many other estimates that organisations routinely make.'

The 'bang for buck' metrics (B) based on the overarching utility of the evaluation; are the motive (M), the design of the evaluation (D) the potential impact (I) and the percentage of recommendations implemented (R). These variables multiplied will give the value (V) of the evaluation. To determine the benefit in relation to the resources expended, it is divided by the total cost of the evaluation (C)

The formula to calculate the ‘bang for buck’ metric is as follows:

$$B = V / C$$

Where $V = M \times I \times R \times D$

Modelling of this formula against numerous evaluation scenarios has enabled a table to be developed to measure the relative effectiveness for the B metric.

(B) metric	Relative effectiveness
>4.0	Optimum
2.0-4.0	Highly effective
1.0 – 2.0	Effective.
<1.0	Ineffective

Table 2 ‘Bang for buck’ metric and relative effectiveness

The formula is limited to comparing evaluations of comparatively equal value and scope. For large projects which require significant resources (eg. Longitudinal evaluations), while a ‘bang for buck’ metric can be assigned, it may be difficult to benchmark against other studies to determine relative value. The ‘bang for buck’ metric is therefore suited for relatively small and simple evaluation projects where more extensive metrics and analysis, such as the Stufflebeam’s (1999) Meta-evaluation Model, do not apply.

Case Studies to test the ‘Bang for Buck’ Formula

Air Force is apt at conducting its own evaluation of its training courses. In some areas, due to staffing limitations, the ability to conduct evaluations of training courses has diminished. In order to continue an evaluation program, these areas initiated a program of outsourcing evaluations. These outsourced evaluations will be the case studies to test the ‘bang for buck’ metrics, as they are easily quantifiable to calculate costs and represent a variety of approaches to evaluation. In effect, their variety ensures they are ideal to test the limitations and scope of the ‘bang for buck’ formula.

Case Study One: Aircraft Trade Course

Evaluation Motive. The scope of the evaluation was to evaluate the course to obtain data and information that would facilitate informed and appropriate decisions as to the type of changes that may be required within the course curriculum to meet organisational requirements. This was basically process improvement at level four motive.

Design Analysis. This evaluation attempted to identify stakeholder requirements, however in reality stakeholders were incorrectly identified as they did not have the authority to make decision within the system. The focus was identified, however there was no further delineation of evaluation questions. Indicators were not identified and measurement tools were basic. The criteria to determine the population sampling was not specified. Analysis and reporting was thorough. No post-evaluation monitoring was evident. Design covered 6 of the Air Force design components.

Potential Impact The study concluded that the majority of the respondents agreed that the training was necessary and satisfied that it met workplace requirements. Impact level of 4.

% of recommendations Implemented. Around 20% of the recommendations were outside of the decision authority of the evaluation commissioner. Approximately 12 recommendations were implemented. Recommendation metric is 48%.

Cost. Total costs of \$41 885 (inclusive \$26 885 contractor cost)

Case Study Two: Supervisors Course

Evaluation Motive. The scope of the evaluation was to evaluate the course to obtain data and information on the effectiveness of the course meeting workplace requirements. This was basically process improvement at level four motive.

Design Analysis. Consideration of stakeholder requirements (in reality stakeholders were incorrectly identified as part of the training system and not the employment system). Data sources were identified through random sampling of graduates over two years. This data could have been ascertained through the less expensive means of internal evaluation. Design is therefore assessed as low in value. No post course monitoring of the evaluation recommendations was evident. Design covered 6 of the Air Force design components.

Impact of Evaluation. Information on assessments, course administration and learning strategy made up around 3/4 of the survey. This type of information can be derived from student feedback evaluations more efficiently than a commissioned evaluation. Impact is low; level 2.

% of recommendations implemented. 27 recommendations were recorded in the report with 60% implemented.

Cost. Total costs of \$54 696 (inclusive of \$39 696 contractor costs)

Case Study Three: Medical Related Course

Evaluation Motive. The stated aim of this evaluation was to gather data about the nature of the course, its value to graduates and supervisors in the workplace. The motive level is 2.

Design. Evaluation questions were included. Data sources were total population. No indication how results would be managed. Problems with sourcing data were never planned for. No monitoring of results. Design covered 7 of the Air Force design components.

Potential Impact. Three recommendations were reported. Recommendation 1 was to validate the data within the report. Recommendation 2 was a non-specific rationalisation task. Recommendation 3 was for broad topic inclusions. Consideration recommendation only. Highest potential impact level is 3.

% of recommendations implemented. Due to the obscure phrasing of the recommendations and non-specificity, only around 40% of the recommendations could be implemented.

Cost. Total cost of \$25 900 (including \$16 900 contractor costs)

Analysis of Results

In terms of 'bang for buck', none of the evaluations deliver optimum results to substantiate the significant investment. Case Studies 2 and 3 resulted in an ineffective use of resources to obtain the evaluation results presented, while Case Study 1, represented marginal value for expenditure. The major problems as exemplified in table 3, have included: deficient design in conforming with basic organisational evaluation design, the lower level of motive for the evaluation, low level of implementation/endorsement of evaluation results and the low overall impact of the evaluation on the organisation. Wholey's (2004) assertion that evaluation can be costly and yet can yield little useful information is confirmed by these studies.

Proj type	(C) Cost	(M) Evaluation Motive	(D)esign	(I) Impact of eval findings on program	(R)% recom'ns implemented	(B) 'Bang for Buck' metric	Notes
TRG Cse 1	42	4	6	4	.65	46.10	1.26
TRG Cse 2	55	4	6	2	.55	52.8	0.96
TRG Cse 3	26	2	7	3	.40	18.8	0.64

Table 3 Summary of results of case studies

Towards an Optimum Evaluation Framework

The evaluation case studies applying the 'bang for buck' formula highlight the inefficiencies and ineffectiveness inherent within evaluations. To move towards an optimum evaluation framework a number of steps can be considered by the evaluator. Firstly, the resources input must match the motive and impact. If the utility is aligned to lower level motives and impacts, such as compliance, then the level of data collection, collation and reporting can be minimised in order to save costs. For compliance evaluations savings could be made through more emphasis on level one reaction data at the conclusion of the program, and reducing the scope of the data collection (eg. focus groups and random sampling of extensive data collection and analysis). Secondly, the evaluation design must be robust and valid. An ineffective design can invalidate the data resulting in limited utility. To achieve higher levels of utility, the organisation must invest in the design of the evaluation. Finally, the recommendations must be presented within the political, economic, social and structural framework of the organisation. A low rate of implementation/endorsement of the recommendations is symptomatic of organisational misalignment and represents a waste of resources. If the motive level is comparatively high, the recommendations should reflect this level.

The 'bang for the buck' formula is an application that proved an effective tool in the Air Force training environment in determining the relative value for outsourced evaluations. As the formula conforms to the generic principles within evaluation theory and practice, it could have broader application across organisational evaluation contexts. Any application outside the training environment is yet to be tested, however as few organisations have the resources to waste on ineffective evaluations; working within a framework to optimise 'bang for the buck' can only assist in maximising the investment.

REFERENCES

- Bemelmans-Videc, M. L., Eriksen, B. and Goldenberg, E.N. (1994). 'Facilitating Organizational Learning: Human Resource Management and Program Evaluation' in Leeuw, F.L., Rist, R.C. and Sonnichsen, R.C., (eds) Can Governments Learn? Comparative Perspectives on Evaluation and Organizational Learning pp. 145-188, New Brunswick, N.J., Transaction Publishers.
- Chelimsky, E. (2006). 'The Purposes of Evaluation in a Democratic Society' in Shaw, I., Greene, C. (eds) and Mark, M. The SAGE Handbook of Evaluation pp. 33-55. London, SAGE Publications Ltd.
- Datta, L. (1997) 'Multimethod Evaluations: Using case studies together with other methods' in Chelimsky, E. and Shadish., W (1997). Evaluation for the 21st Century: a Handbook. pp. 344 – 359, Thousand Oaks, Ca., Sage Publications
- Hendricks, M and Papagiannis, M. (1982) 'Oral policy briefings' in Smith (ed.) Communication Strategies in Evaluation pp. 249-258 Beverly Hills, CA, Sage.
- Henry, G. (2000). Why not use ? *New Directions for Evaluation*, (88), 85-98
- Hipwell, W. (2005) "Value Creation and Measurement in Learning Environments." *Chief Learning Officer: 2005 Business Intelligence Industry Report Volume*, DOI:
- King, J. A. (2004). 'The Roots of Participatory Evaluation' in Alkin, M.C. (ed) Evaluation Roots: Tracing Theorists' Views and Influences. pp. 331 - 342, Thousand Oaks, Ca., Sage Publications, Inc.
- Nevo, D. (2006) 'Evaluation in Education' in Shaw, I., Greene, C. (eds) and Mark, M. The SAGE Handbook of Evaluation pp. 439-460 London, SAGE Publications Ltd.
- Owen, J. M. (2003). "Evaluation Culture: a definition and analysis of its development within organisations." *Evaluation Journal of Australasia* **3 (new series)**(1): 43-47
- Patton, M.Q. (1990) Qualitative Evaluation and Research Methods Newbury Park, Ca., SAGE Publications Ltd.
- Patton, M. Q. (2004). 'Utilization-Focussed Evaluation' in Alkin, M.C. (ed) Evaluation Roots: Tracing Theorists' Views and Influences. pp. 276 - 292. Thousand Oaks, Ca., Sage Publications, inc.
- Phillips, J. J. (1998). 'ROI: The Search for Best Practices' in Kirkpatrick, D. *Another Look at Evaluating Training Programs* pp. 210-215. Alexandria, Va., ASTD
- Rist, R. (1994). 'The Preconditions for Learning: Lessons from the Public Sector' in Leeuw, F.L., Rist, R.C. and Sonnichsen, R.C., (eds) Can Governments Learn? Comparative Perspectives on Evaluation and Organizational Learning pp. 146-189. New Brunswick, N.J., Transaction Publishers.
- Rogers, P. and Williams., B. (2006). 'Evaluation for Practice improvement and Organizational Learning' in Shaw, I., Greene, C. (eds) and Mark, M. The SAGE Handbook of Evaluation pp. 76 - 98. London, SAGE Publications Ltd
- Rossi, P. H., Lipsey, M.W. and Freeman, H.E. (2004). Evaluation: A Systematic Approach. Thousand Oaks, Ca., Sage Publications, Inc.
- Schadlock, R. L. and Thornton., C.V. (1988). Program Evaluation: A field Guide for Administrators. New York, Plenum Publishing Corporation.
- Scmidt, W. (1998). 'Cost benefit Analysis Techniques for Training Investments in Kirkpatrick, D (Ed) *Another Look at Evaluating Training Programs* pp. 185-188. Alexandria, VA, ASTD.
- Stake, R. and Schwandt, T. (2006) 'On discerning quality in evaluation' in Shaw, I.F., Green, J.C. and Mark, M.M. Handbook of Evaluation, pp. 404-418 London, SAGE
- Stufflebeam, Daniel L. 'The metaevaluation imperative'. *American Journal of Evaluation*, 22 (2), 183-210. Retrieved July 22, 2008 from EBSCO host database.
- Walker, R. and Wiseman, M. (2006) 'Managing Evaluations' in Shaw, I., Greene, C. (eds) and Mark, M. The SAGE Handbook of Evaluation pp. 360 - 383. London, SAGE Publications Ltd
- Wholey, J. S. (2004). 'Evaluation, Performance, and Decision Making' in Alkin, M.C. (ed) Evaluation Roots: Tracing Theorists' Views and Influences. pp. 267-275. Thousand Oaks, Ca., Sage Publications, Inc.