

EVALUATING MANAGEMENT DEVELOPMENT

DR. JOHN KENWORTHY



CELSIM

www.celsim.com

EVALUATION

BACKGROUND

There are several issues involved in evaluating management development, more so when the evaluation also considers the method of management development that this paper explores. Through a review of the literature, this paper will address the high level issues and establish the basis for a suitable model of evaluation that will measure the effectiveness of the method(s) employed in a managerial development intervention or event.

The paper commences by reviewing **why** we should evaluate; **what** should be evaluated; and lastly **how** to evaluate. The paper then considers suitable models and methods of evaluation to measure the effectiveness of a management development intervention and the method employed in undertaking the intervention with particular reference to experiential learning and the use of computer-based simulations during a training intervention.

WHY EVALUATE

A number of authors consider the need and reasons for evaluation though all tend to fall into four broad categories identified by Mark Easterby-Smith (Easterby-Smith, 1994). He notes four general purposes of evaluation (P14):

1. *Proving*. the worth and impact of training. Designed to demonstrate conclusively that something has happened as a result of training or developmental activities.
2. *Improving*. A formative purpose to explicitly discover what improvements to a training programme are needed
3. *Learning*. Where the evaluation itself is or becomes an integral part of the learning of a training programme
4. *Controlling*. Quality aspects in the broadest sense, both in terms of quality of content and delivery to established standards.

More recent literature uses different terminology which can be placed into these four broad categories:

Russell (1999) does not include such a *Learning* purpose explicitly but suggests that the evaluation of performance interventions produces the following benefits:

- Determines if performance gaps still exist or have changed (Proving)
- Determines whether the performance intervention achieved its goals (Controlling)
- Assesses the value of the performance intervention against organizational results (Proving)
- Identifies future changes or improvements to the intervention (Improving)

According to Russell, whichever evaluation model is chosen to follow “the evaluation should focus on how effective the performance intervention was at meeting the organization’s goals”.

Russ-Eft and Preskill (2001) review evaluation across the entire organisation, not just in the training or development arena and cite 6 distinct reasons to evaluate:

Evaluation:

- Ensures quality (Controlling)
- Contributes to increased organisation members' knowledge (Learning)
- Helps prioritise resources (Improving)
- Helps plan and deliver organisational initiatives (Improving)
- Helps organisation members be accountable (Controlling)
- Findings can help convince others of the need or effectiveness of various organisational initiatives (Proving)

They include a seventh reason to gain experience in evaluation as it is 'fast becoming a marketable skill'.

Increasingly in the business and Human Resources media and the professional bodies representing training and development professionals (for example the American Society for Training and Development, ASTD), there is a call for 'better' evaluation of training and development intervention. The basic principle driving this is for training to demonstrate its worth to organisations – whether that be its' attributable Return on Investment (ROI) or its value in improving performance of individuals (such as productivity gains or reduced accidents) or of the organisation (such as more efficient use of resources or demonstrable improvement in quality). Essentially, training and development costs time and money and needs to be shown to be worthwhile.

WHAT TO EVALUATE

Exactly what is to be measured as part of the evaluation is an especially problematic area. Aspects of behaviour or reaction that are relatively easy to measure are usually trivial. Measuring changes in behaviour, for example, may require the observations to be reduced to simpler and more trivial characteristics. Alternatively, these characteristics may be assessed by the individual's subordinates – however, the purist would no doubt claim that such holistic judgements are of dubious validity. The problem is that the general requirement for quantitative measurement tends to produce a trivialization in the focus of the evaluation.

According to Bedingham (1997), ultimately "the only criteria that make sense are those which are related to on-the-job behaviour change". Bedingham advocates the use of research based 360° questionnaires that objectively measure competency sets and skills applicable to most organisations, functions or disciplines and making the results of the feedback taken immediately prior to the event, available to trainees during their training – thus allowing individuals to easily see how they actually do something and the relevance of the training. Thus, they can then start transferring the learned skills immediately on return to the workplace.

MANAGERIAL EFFECTIVENESS AND COMPETENCY

David McClelland is often cited as the father of the modern competency movement. In 1973, he challenged the then orthodoxy of academic aptitude and knowledge content tests, as being unable to predict performance or success in life and as being biased against minorities and women (Young, 2002). Identified through patterns of behaviour, competencies are characteristics of people that differentiate performance in a specific job or role (Kelner, 2001, McClelland, 1973). Competencies distinguish well between roles at the same level in different functions and also between roles at different levels (even in the same function) often by the number of competencies required to define the role. A competency model for a Middle Manager, is usually defined within

ten to twelve competencies, of those two are relatively unique to a given role. Kelner suggest that competency models for senior executives require fifteen to eighteen competencies, up to half of which were unique to the model (Kelner, 2001).

Kelner (2001), cites a 1996 unpublished paper by the late David McClelland where he performed a meta-analysis of executives assessed on competencies, where McClelland discovered that only eight competencies could consistently predict performance in any executive with 80 percent accuracy.

The first scientifically valid set of scaled competencies – competencies that have sets of behaviours ordered into levels of sophistication or complexity – were developed by Hay/McBer (Spencer and Spencer, 1993). The competencies found to be the most critical for effective managers include:

- Achievement Orientation
- Developing Others
- Directiveness
- Impact and Influence
- Interpersonal Understanding
- Organisational Awareness
- Team Leadership

This set of characteristics, or individual competencies, that a manager brings to the job need to match well to the job or additional effort may be necessary to carry out the job or the manager may not be able to use certain managerial styles effectively. These are in turn affected by the organisational climate and the actual requirement of the job.

Managerial effectiveness is the combination of these four critical factors, Organisational Climate, Managerial Styles, Job Requirements and the Individual Competencies that the manager brings to the job. Reddin (1970) points out that 'managerial effectiveness' is not a quality but a statement about the relationship between his behaviour and some task situation. According to Reddin, it is therefore not possible to discover as fact the qualities of effectiveness which would then be of self-evident value.

If, however the Hay/McBer competencies are able to predict performance in any executive to 80 percent accuracy and these competencies are 'trainable' then any training programme designed to develop managerial effectiveness in any role can be evaluated by means of assessing the changes in behaviour of the participant that demonstrates the competency.

LEARNING OUTCOMES

Most researchers conceptualize learning as a multidimensional construct. There is considerable commonality across different attempts to classify types of learning outcomes. (Kraiger et al., 1993) synthesised the work of (Gagne, 1984), (Anderson, 1982) and others proposing broad based categories of learning outcomes: skill-based, cognitive and affective outcomes.

Skill-based learning outcomes address technical or motor skills. A number of game-based instructional programmes have been used for practice and drill of technical skills. For example, (Gopher et al., 1994) investigated the effectiveness of the use of an aviation computer game by military trainees on subsequent test flights.

Cognitive learning outcomes include three subcategories of:

- Declarative knowledge. The learner is typically required to reproduce or recognise some item of information. For example, (White, 1984) demonstrated that students who played a computer game focusing on Newtonian principles were able to more accurately answer questions on motion and force than those who did not play the game.
- Procedural knowledge. Requires a demonstration of the ability to apply knowledge, general rules, or skills to a specific case. For example, (Whitehall and McDonald, 1993) found that students using a variable payoff electronics game achieved higher scores on electronics troubleshooting tasks than students who received standard drill and practice.
- Strategic knowledge. Requires application of learned principles to different contexts or deriving new principles for general or novel situations (referred to by others, as Constructivist Learning (Dede, 1997)), (Wood and Stewart, 1987), for example, found that use of a computer game to improve practical reasoning skills of students led to improvements in critical thinking.

Affective learning outcomes refers to attitudes. Includes feelings of confidence, self-efficacy, attitudes, preferences and dispositions. Some research has shown that games can influence attitudes, for example (Wiebe and Martin, 1994) and (Thomas et al., 1997).

GENERAL APPROACHES TO EVALUATION 'SCHOOLS'

Easterby-Smith groups major approaches and classifies within two dimensions, the scientific-constructivist dimension, and the research-pragmatic dimension.

The Scientific-constructivist dimension represents the distinct paradigms – according to Filstead (1979), these paradigms represent distinct, largely incompatible, ways of seeing and understanding the world, yet In practice, most evaluations contain elements of each point of view (Easterby-Smith, 1994). The scientific approach favours the use of quantitative techniques involving the attempt to operationalise all variables in measurable terms – normally analysed by statistical techniques in order to establish absolute criteria.

The **Scientific approach** contrasts greatly with 'constructivist' or 'naturalistic' methods which emphasise the collection of different views of various stakeholders before data collection begins. The process continues with reviewing largely, but not exclusively, qualitative data along with the view of various stakeholders.

Research-Pragmatic dimension

This dimension represents the contrasting styles of how the evaluation is conducted. Easterby-Smith (1980) described the two extremes as *Research* and *Pragmatic* representing the Research and Pragmatic styles respectively.

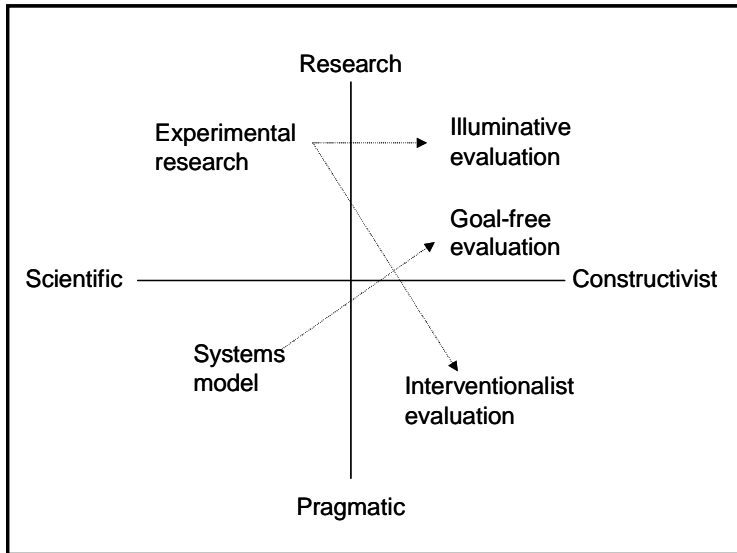
Research styles stress the importance of rigorous procedures, that the direction and emphasis of the evaluation study should be guided by theoretical considerations and that these considerations are aimed at producing enduring generalizations, and knowledge about the learning and developmental process involved. The researcher in this instance should be independent and maintain an objective view of the courses under investigation without becoming personally involved.

The pragmatic approach, in contrast, emphasizes the reduction of data collection and other time-consuming aspects of the evaluation study to the minimum possible. As pointed out in (Easterby-Smith et al., 1991), when working within companies, the researcher is dependent on the cooperation and given time of managers and other informants – who may and will have other more important priorities than the researcher's study.

Easterby-Smith combines the dimensions into a useful matrix. The arrows show the influence of one 'school' on the development of the linked 'school'.

(Figure 1)

Figure 1. Models and 'schools of thought' in evaluation



EXPERIMENTAL RESEARCH

Experimental research has its roots in traditional social science research methodology. (Campbell and Stanley, 1966), (Campbell et al., 1970), (Kirkpatrick, 1959/60), (Hesseling, 1966) are cited in (Easterby-Smith, 1994) as the best known representatives of this 'school'. The emphasis in experimental research is:

1. Determining the effects of training and other forms of development
2. Demonstrating that any observed changes in behaviour or state can be attributed to the training, or treatment, that was provided.

There is an emphasis on the theoretical considerations, preordinate designs and quantitative measurement and emphasis on comparing the effects of different treatments. In a training evaluation study this would involve at least two groups being evaluated before and after a training intervention. One group receives a training 'treatment' whilst the other group does not. The evaluation would measure the differences between the groups in specific quantifiable aspects related to their work.

KIRKPATRICK MODEL

Donald Kirkpatrick created the most familiar evaluation taxonomy of a four step approach to evaluation (Kirkpatrick, 1959/60) – now referred to as a model of four levels of evaluation (Kirkpatrick, 1994). It is one of the most widely accepted and implemented models used to evaluate training interventions. Kirkpatrick's four levels measure the following:

1. Reaction to the intervention
2. Learning attributed to the intervention
3. Application of behaviour changes on the job
4. Business results realised by the organisation

Russ-Eft and Preskill (2001) note that the ubiquity of Kirkpatrick's model stems from its simplicity and understandability – having reviewed 57 journal articles in the training, performance, and psychology literature that discussed training evaluation models, Russ-Eft and Preskill found that 44 (77%) of these included Kirkpatrick's model. They also note that only in recent years (1996) that several alternative models have been developed.

In an article written in 1977, Donald Kirkpatrick considered how the evaluation at his four levels provided evidence or proof of training effectiveness. Proof of effectiveness requires an experimental design using a control group to eliminate possible factors affecting the measurement of outcomes from a training program. Without such a design, i.e., evaluating only the changes in a group participating in a training program, can provide evidence of training effectiveness but not proof (Kirkpatrick, 1998).

BRINKERHOFF MODEL

Brinkerhoff's model has its roots in evaluating training and HRD interventions. Brinkerhoff's cyclical model consists of six stages grouped into the following four stages of performance intervention:

- Performance Analysis
- Design
- Implementation
- Evaluation

Brinkerhoff's (1988, 1989) model addresses the need for evaluation throughout the entire human performance intervention process.

The six stages:


1. Goal setting – identify business results and performance needs and determine if the problem is worth addressing
2. Program design – evaluation of all types of interventions that may be appropriate

3. Program implementation – Evaluates the implementation and addresses the success of the implementation
4. Immediate outcomes – focuses on learning that takes place during the intervention
5. Intermediate outcomes – focuses on the after-effects of the intervention some time following the intervention
6. Impacts and worth – how the intervention has impacted the organization, the desired business results and whether it has addressed the original performance need or gap.

HOLTON'S HRD EVALUATION RESEARCH AND MEASUREMENT MODEL

Holton (1996) identified three outcomes of training (learning, individual performance, and organisational results) that are affected by primary and secondary influences. Learning is affected by trainee's reactions, their cognitive ability, and their motivation to learn, The outcome of individual performance is influenced by motivation to transfer their learning, the programme's design, and the condition of training transfer. Organisational results are affected by the expectations for return on investment, the organisation's goals, and external events and factors. Holton's model bears great similarity to Kirkpatrick's Levels 2, 3 and 4.

The model is more testable than others in that it is the only one that identifies specific variables that affect training's impact through identification of various objects, relationships, influencing factors, predictions and hypotheses. Russ-Eft view Holton's model as being related to a theory-driven approach to evaluation (Russ-Eft and Preskill, 2001).



ISSUES WITH EVALUATION APPROACH

EXPERIMENTAL APPROACH

There are a number of reasons why this approach may not work as well as it might, particularly with management training where sample sizes are limited and especially where training and development activities are secondary to the main objectives of the organization. Easterby-Smith (1994) discusses four main reasons:

SAMPLE SIZES

Using statistical techniques, essential to experimental research, need to be large to discover statistical significances when evaluating management training and development. This is particularly difficult in this field when group sizes are often less than ten and rarely greater than 30.

CONTROL GROUPS

There are special problems in achieving genuine control groups. In one study (Easterby-Smith and Ashton, 1975), the selection of the group to receive training were 'closer' to their bosses than those who were not selected – rather than more objective or even random selection criteria.

Further, does 'no training' have a negligible effect on the 'control group'? Indeed Guba and Lincoln (1989) now include those who are left out as one of the main groups of stakeholders.

CAUSALITY

The intention of experimental design is primarily to demonstrate causality between the training intervention and any subsequent outcomes. However, it is often hard to isolate the intervention from other influences on the manager's behaviour. It may be possible to reduce the 'clutter' of other influences for training interventions that have a specific, identifiable skills focus, interventions of a more complex nature may be subject to myriad external influences between evaluations.

TIME AND EFFORT

Kirkpatrick identifies that an experimental design to 'prove' that training is effective requires additional steps including the evaluation of a control group. Particularly with regard to his Level 3 (Behaviour) and Level 4 (Results), ensuring that any post-test evaluation is undertaken at a time after the training event long enough for the participant to have had an opportunity to change behaviour, or for results to be realisable (Kirkpatrick, 1998).

ILLUMINATIVE EVALUATION

In response to the problems above with experimental research, evaluators usually try to increase the sample size. The Illuminative evaluation school takes issue with the comparative and quantitative aspects of experimental research. Such a view will tend to concentrate rather on the views of different people in a more qualitative way. However, this school has been noted to be more costly than anticipated (Jenkins et al., 1981), and also this style of evaluation had been rejected by sponsors.

SYSTEMS MODEL

Easterby-Smith (1994) notes three main features of the systems model 'school'. Starting with the *objectives*, an emphasis on identifying the *outcomes* of training and a stress on providing *feedback* about these outcomes to those people involved in providing inputs to the training.

Additionally, *evaluation* is the 'assessment of the total value of a training system, training course or programme in social as well as financial terms'. Critical of the approach Hamblin (1974) suggests that this is restrictive because evaluation has to be related to objectives whilst being over ambitious because the total value of training is evaluated in 'social as well as financial' terms. However, Hamblin's own cycle of evaluation depends heavily on the formulation of objectives, either as a starting point or as a product of the evaluation process.

Another important feature of Hamblin's work is the emphasis on measurement of *outcomes* from training at different levels. It assumes that any training event will, or can, lead to a chain of consequences, each of which may be seen as causing the next consequence. Hamblin stresses that it would be unwise to conclude from an observed change at one of the higher levels of effect that this was due to a particular training intervention, **unless** one has followed the chain of causality through the intervening levels of effect. Should a change in job behaviour, for example, be observed, the constructivist take on this would possibly be to ask the individual for his or her own views of why they were now behaving in a different way and then compare this interpretation with the views of one or two close colleagues or subordinates.

HAMBLIN'S FIVE-LEVEL MODEL

Hamblin (1974), also widely referenced, devised a five-level model similar to Kirkpatrick's. Hamblin adds a fifth level that measures "ultimate value variables" of "human good" (economic outcomes). This also can be viewed as falling into the tradition of the behavioural objectives approach (Russ-Eft and Preskill, 2001).

(Figure 2)

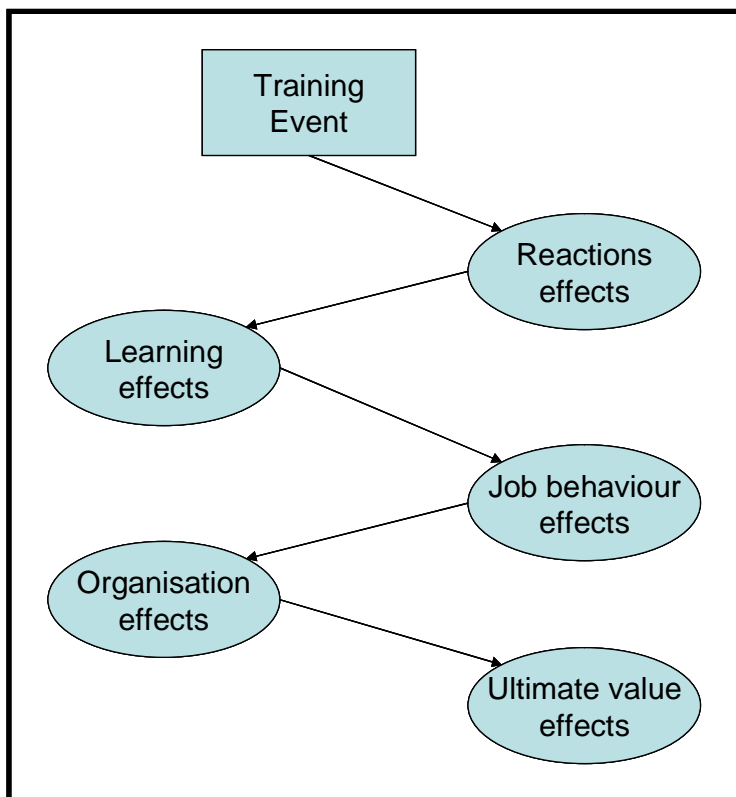


Figure 2. 'Chain of Consequences' for a training event

The third main feature of the systems model is the stress on feedback to trainers and others decision makers in the training process. This features significantly in the work of (Warr et al., 1970) who take a very pragmatic view of evaluation, suggesting that it should be of help to the trainer in making decisions about a particular programme as it is happening. (Rackham, 1973) subsequently makes a further distinction between assisting decisions that can be made about *current* programmes, and feedback that can contribute to decisions about *future* programmes. This, Rackham began to appreciate after attempting to improve the amount of learning achieved in successive training programmes by feeding back to the trainers data about the reactions and learning achieved in earlier programmes. What Rackham noticed was that the process of feedback from one course to the next resulted in clear improvements when the programmes were *non-participative* in nature, but that there were no apparent improvements in programmes that involved a lot of participation.

The idea of feedback as an important aspect of evaluation was developed further by Burgoyne and Singh (1977). They distinguish between evaluation as *feedback* and feedback adding to the body of *knowledge*. The former they saw as providing transient and perishable data relating directly to decision-making, and the latter as generating permanent and enduring knowledge about education and training processes.

Burgoyne and Singh relate evaluative feedback to a range of different decisions about training in the broad sense:

1. Intra-method decisions: about how particular methods are handled, for example 'lectures' may vary from straight delivery to lively debates.
2. Method decisions: for example, whether to employ a lecture, a case study, or a simulation in order to introduce the topic
3. Programme decisions: about the design of a particular programme, whether it should be longer or shorter, more or less structured, taught by insiders or visiting speakers
4. Strategy decisions: about the optimum use of resources, and about the way the training institution might be organized.
5. Policy decisions: about the overall provision of funding and resources, and the role of the institution as a whole, whether for example, a management training college should see itself as an agent for change or as something that oils the wheels of change that are already taking place. (Figure 3)

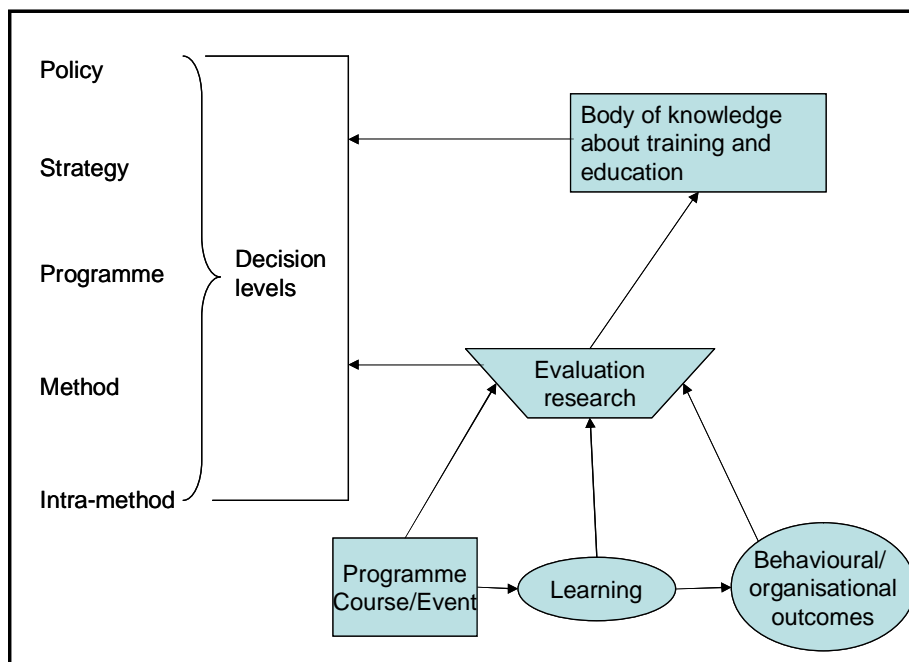


Figure 3. Evaluation of outcomes and link to decision making

The systems model has been widely accepted, especially in the UK, but there are a number of problems and limitations that should be understood. According to Easterby-Smith (1994) the main limitation is that feedback (i.e. data provided from evaluation of what has happened in the past) can only contribute marginally to decisions about what should happen in the future. This is due to the legacy of the past training event and feedback can highlight incremental improvements based on the previous design, but not note when radical change is needed.

The emphasis on outcomes provides a good and logical basis for evaluation but it represents a mechanistic view of learning. In the extreme, this suggests that learning consists of facts and knowledge being *placed* in

people's heads and that this becomes internalized and then gradually incorporated in people's behavioural responses.

The emphasis on starting with *objectives* brings us to a classic critique of the systems approach. Just whose objectives are they? It has been questioned by MacDonald-Ross (1973) whether there is any particular value in specifying formal objectives at all, since among other things, this might place undue restrictions on the learning that could be achieved from a particular educational or training experience.

GOAL FREE EVALUATION

This leads to the next major evaluation 'school': Goal free evaluation. This starts with the assumption that the evaluator should avoid consideration of formal objectives in carrying out his or her work. Scriven (1972) proposed the radical view that the evaluator should take *no* notice of the formal objectives of a programme when carrying out an investigation of it.

Goal free evaluation leans more to the constructive method where the evaluator should avoid discussing or even seeing the published objectives of the programme and discover from participants what happened and what was learned (Scriven, 1972).

INTERVENTIONALIST EVALUATION

According to Easterby-Smith (1994) this approach includes *responsive* evaluation and *utilization focused* evaluation. He cites Stake (1980) contrasts *responsive* evaluation with the preordinate approach of experimental research. The latter requires the design to be clearly specified and determined before evaluation begins, it makes use of 'objective' measures, evaluates these against criteria determined by programme staff, and produces reports in the form of research-type reports. In contrast, responsive evaluation is concerned with programme activities rather than intentions, and takes account of different value perspectives. In addition, Stake stresses the importance of attempting to respond to the audience's requirements for information (contrasting with some goal-free evaluators to distance themselves from some of the principal stakeholders). Stake positions this method more pragmatically by recognizing that 'different styles of evaluation will serve different purposes'. He also recognizes that preordinate evaluations may be preferable to responsive evaluations under certain circumstances.

Guba and Lincoln (1989) take this method further by what they call *responsive constructivist evaluation*. Guba and Lincoln recommend starting with the identification of stakeholders and their concerns, and arranging for these concerns to be exchanged and debated before collection of further data.

Patton (1978) stresses the importance of identifying the motives of key decision makers before deciding what kind of information needs to be collected. Recognising that some stakeholders have more influence than others (a view Guba and Lincoln argue *should* not be the case) but goes further by concentrating on the uses of the subsequent information might be put.

But, Interventionalist evaluation has the danger of being so flexible in its approach because it considers the views of all stakeholders and the use of the subsequent information, that the form adapts and changes to every passing whim and circumstance, and thereby producing results that are weak and inconclusive. It may also be, that the evaluator becomes too close to the programme and stakeholders, something that goal-free evaluators set out to avoid, leading to a reduction of impartiality and credibility.

EXPERIENCE WITH TRAINING PROGRAMME EVALUATION MODELS

Many researchers measuring the effects of training have looked at one or more of the outcomes identified by Kirkpatrick (1959/60, 1994): reactions, learning, behaviour, and results.

Evaluation of Trainee reactions to learning has yielded mixed results (Russ-Eft and Preskill, 2001), yet it is often the only formal evaluation of a training programme and relied upon to assess learning and behaviour change. It may be reasonable to assume that enjoyment is a precursor to learning, and that if a trainee enjoys the training, they are likely to learn but such an assumption is not supported by a meta-analytic study combining the results of several other studies (Alliger et al., 1997).

Kirkpatrick's second level, Learning, is the most commonly used measurement after trainee reactions to assess the impact of training. Studies investigating the relationship between learning and work behaviour have also shown mixed results and offer little concrete evidence to support the notion that increased learning from a training programme relates to organisational performance. (Collins, In Press) cited in (Russ-Eft and Preskill, 2001).

Training transfer is defined as applying the knowledge, skills, and attitudes acquired during training to the work setting. Most organisations are genuinely interested in this aspect of the effectiveness of training events and programmes – yet the paucity of research dedicated to transfer of training contradicts the importance of the issues. Some research in this area has focussed on comparison of alternative conditions to training transfer. A typical design of such research compares groups who receive different training methods and/or a 'control group' that receives no training. Such studies do not all use the same research design and the results are inconsistent. However, the research does indicate that some form of post-training transfer strategy facilitates training transfer (Russ-Eft and Preskill, 2001).

Evaluating Results of training in terms of business results, financial results and return on investment (ROI) is much discussed in the popular literature – most offering anecdotal evidence or conjecture about the necessity of evaluating training's return on investment and methods that trainers might use to implement such an evaluation. Solid research on this topic is not, however, so voluminous. Mosier (1990) proposes a number of capital budgeting techniques that could be applied to evaluating training whilst recognising that there are common reasons why such financial analyses are rarely conducted or reported:

- It is difficult to quantify or establish a monetary value for many aspects of training
- No usable cost-benefit tool is readily available
- The time lag between training and results is problematic
- Training managers and trainers are not familiar with financial models.

Little progress appears to have been made in this area since Mosier wrote.

CHOOSING AN EVALUATION STYLE

EXPERIENCE WITH TRAINING PROGRAMME EVALUATION MODELS

Reviewing the history and development of training evaluation research shows that there are many variables that ultimately affect how trainees learn and transfer their learning in the workplace. Russ-Eft and Preskill suggest that a comprehensive evaluation of learning, performance, and change would include the representation of a considerable number of variables (Russ-Eft and Preskill, 2001). Such an approach, whilst laudable in terms of purist academic research, is likely to cause another problem, that of collecting data to demonstrate the affects and effects of so many independent variables and factors. Thus, we need to recognise that there is a trade off between the cost and duration of a research design and increasing the quality of the information which it generates (Warr et al., 1970).

Hamblin (1974) points out that a considerable amount of evaluation research has been done. This research has been carried out with a great variety of focal theories, usually looking for consistent relationships between educational methods and learning outcomes, using a variety of observational methods but with a fairly consistent and stable background theory. However, the underlying theory has been taken from behaviourist psychology summed up as the 'patient' - here the essential view is that the subject (patient) does (behaviour or response) is a consequence of what has been done to him (treatment or stimulus).

Another approach according to Burgoyne (Burgoyne and Cooper, 1975) which researchers appear to take to avoid confronting value issues is to hold that all value questions can ultimately be reduced to questions of fact. This usually takes the form of regarding the quality of 'managerial effectiveness' as a personal quality which is potentially objectively measurable, and therefore a quality, the possession of which could be assessed as a matter of fact. In practice this approach has appeared elusive. Seashore et al, (Seashore et al., 1960) felt that the existence of the concept of 'managerial effectiveness' as a unitary trait could be confirmed, if they found high intercorrelations between five important aspects of managerial behaviour: high overall performance ratings, high productivity, low accident record, low absenteeism and few errors

There are no forced rules about the style of evaluation used for a particular evaluative purpose. However, (Easterby-Smith, 1994) suggests that studies aimed at fulfilling the purpose of *proving* will tend to be located towards the 'research' end of the dimension, and studies aimed at *improving* will tend to be located near the 'pragmatic' end. On the methodological dimension there may be more concern with *proving* at the 'scientific' end, and *learning* at the 'constructivist' end. (Figure 4)

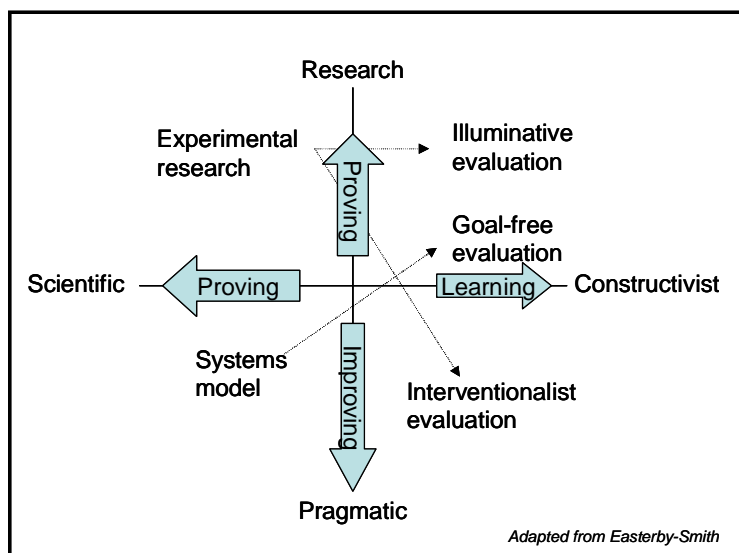


Figure 4. Use of evaluation style matrix

ISSUES OF EVALUATION FOR SIMULATIONS

Margaret Gredler (1996) notes that a major design weakness of most studies evaluating simulation based training and development is that they are compared to regular classroom instruction. However, the instructional goals for each can differ. Similarly, many studies show measurement problems in the nature of the post-tests used.

The problems highlighted above regarding the choice of evaluation methodology are further compounded by the lack of well-designed research studies in the development and use of games and simulations (Gredler, 1996) – much of the published literature consists of anecdotal reports and testimonials providing sketchy descriptions of the game or simulation and report only on perceived student reactions.

Most of the research, as noted by Pierfy (1977) is flawed by basic weaknesses in both design and measurement. Some studies implemented games or simulations that were brief treatments of less than 40 minutes and assessed effects weeks later. Intervening instruction in these cases, however, contaminated the results.

DO SIMULATIONS REQUIRE A DIFFERENT ASSESSMENT OF EFFECTIVENESS?

Mckenna (1996) cites many references to papers criticising the evaluation of learning effectiveness using Interactive Multimedia and many researchers ask for more research (Clark and Craig, 1992) (Reeves, 1993) to quantitatively evaluate the real benefits of simulations and games-based learning.

The question whether Simulation based learning requires a new and different assessment techniques beyond those in use today remains relatively unexplored. Although there have been some attempts at constructing theoretical frameworks for the evaluation of virtual worlds (Rose, 1995), (Whitelock et al., 1996), very few working examples or reports on the practical use of these frameworks exists.

Whitelock et al. (1996) argue that effective evaluation methods need to be established to discover if conceptual learning takes place in virtual environments. In practice, however, the assessment of virtual environments and simulations has been focused primarily on its usefulness for training and less on its efficacy for supporting learning especially in domains with a high conceptual and social content.

Hodges (1998) suggests that simulations (especially Virtual Reality) prove most valuable in job training where hands-on practice is essential but actual equipment cannot often be used. Greeno et al. (1993) suggest that knowledge of how to perform a task is embedded in the contextual environment. Druckman and Bjork (1994) suggest that only when a task is learned in its intended performance situations can the learned skills be used in those situations.

Since the early days of simulation and gaming as a method to teach, there have been calls for hard evidence that support the teaching effectiveness of simulations (Anderson and Lawton, 1997). Their paper states that

“despite the extensive literature, it remains difficult, if not impossible, to support objectively even the most fundamental claims for the efficacy of games as a teaching pedagogy. There is little relatively little hard evidence that simulations produce learning or that they are superior to other methodologies.”

They go on to review the reasons as being traceable to the selection of dependent variables and to the lack of rigour with which investigations have been conducted.

HOW TO EVALUATE EFFECTIVENESS OF SIMULATIONS

One of the major problems of simulations is how to “evaluate the training effectiveness [of a simulation]” (Feinstein and Cannon, 2002) citing (Hays and Singer, 1989). Although for more than 40 years, researchers have lauded the benefits of simulation (Wolfe and Crookall, 1998), very few of these claims are supported with substantial research (Miles et al., 1986) (Butler et al., 1988) (Wolfe, 1981, Wolfe, 1985, Wolfe, 1990)

Many of the above cited authors attribute the lack of progress in simulation evaluation to poorly designed studies and the difficulties inherent in creating an acceptable methodology of evaluation.

There are a number of empirical studies that have examined the effects of game-based instructional programs on learning. For example both Ricci, et al. (1996) and Whitehall and McDonald (1993) found that instruction incorporating game features led to improved learning. The rationale for these positive results varied, given the different factors examined in these studies. Whitehall and McDonald argued that the incorporation of a variable payoff schedule within the simulation led to increased risk taking among students, resulting in greater persistence and improved performance. Ricci, et al. proposed that instruction incorporating game features enhanced student motivation, leading to greater attention to training content and greater retention.

However, although anecdotal evidence suggests that students seem to prefer games over other, more traditional methods of instruction, review have reported mixed results regarding the training effectiveness of games and simulations. Pierfy (1977) evaluated the results of 22 simulation-based training game effectiveness studies to determine patterns of training effectiveness. 21 studies are reported on having comparable assessments. Three of the studies reported results favouring the effectiveness of games, three studies reported results favouring the effectiveness of conventional teaching. The remaining 15 studies reported no significant differences. 11 of the studies tested for retention of learning, eight of these indicated that retention was superior for games, the remaining 3 yielded no significant result. 7 of 8 of the studies assessing student preference for training games over classroom instruction reported greater interest in simulation game activities over conventional teaching methods. More recently, Druckman (1995) concluded that games seem to be effective in enhancing motivation and increasing student interest in subject matter, yet the extent to which this translates into more effective learning is less clear.

THE PROBLEM OF SIMULATION VALIDATION

Specifically with regard to simulation based training programmes (but also applying to all training delivery methods perhaps with different terminology), this section reviews the literature on simulation evaluation developing a coherent framework for pursuing the evaluation problem. Three prominent constructs appear in the literature: fidelity, verification, and validation. Validation of a simulation as a method of teaching and that the simulation programme as a training intervention produces (or helps to produce) learning and transfer of learning are the important criteria (Feinstein and Cannon, 2002) yet fidelity and verification are easier to evaluate (but not necessarily to measure objectively) and often distract evaluators from the more tricky issue of validation.

SIMULATION FIDELITY

Fidelity is the level of realism that a simulation presents to a learner. Hays and Singer (1989) describe fidelity as “how similar a training situation must be, relative to the operational situation, in order to train most efficiently”. Fidelity focuses on the equipment that is used to simulate a particular learning environment.

In more sophisticated (technologically) simulations that use virtual reality for example, the construct of fidelity has an additional dimension, that of presence (the degree to which an individual believes that they are immersed within the virtual world or simulation) (Bailey and Witmer, 1994) (Witmer and Singer, 1994) (Dede, 1997) (Salzman et al., 1999).

The degree of fidelity or presence in a learning environment is a difficult element to measure (Witmer and Singer, 1994). Much research during the 60's and 70's studied the relationship between fidelity and its effects on training and education. These studies according to Feinstein and Cannon found that a higher level of fidelity does not translate into more effective training or enhanced learning (Feinstein and Cannon, 2002) In fact, it may be that lower levels of fidelity but with effective 'human virtual environment interface' (navigational simplicity) and a significant degree of presence can assist in trainees acquiring knowledge or skills within the simulation (Salzman et al., 1999) (Stanney et al., 1998) (Gagne, 1984) (Alessi, 1988).

SIMULATION VERIFICATION

Verification is the process of assessing that a model is operating as intended. Verification "is a process designed to see if we have built the model right" (Pegden et al., 1995). During the process, simulation developers need to test and debug errors in the simulation (now usually software errors) through a series of alpha and beta tests under different conditions, verifying that the model works as intended. Often, developers are distracted by this process producing what appear to be brilliant models that work 'correctly' but with no appreciation of the educational effect and hence their validity. In this sense, verification can be a trap, notwithstanding its critical status as a necessary condition of validity (Feinstein and Cannon, 2002).

VALIDATION

Building on the work of Feinstein and Cannon, Cannon and Burns and others, there are three main questions beyond those of the general issues of evaluation noted above. (Figure 5)

1. Are we measuring the 'right' thing? Validity of the constructs
2. Does the simulation provide the opportunity to learn the 'right' thing? Verification of the simulation and the appropriate fidelity for the content and audience.
3. Does the method of using a simulation deliver the learning? Validity as a method of developing.

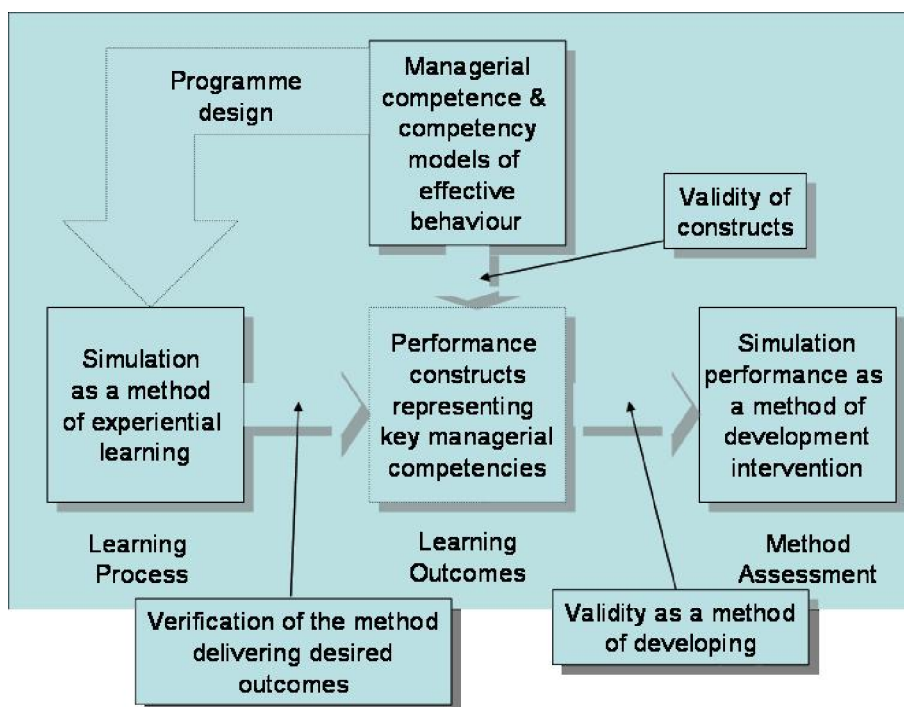


Figure 5. Three Faces of Simulation Evaluation - adapted from Anderson, Cannon, Malik, and Thayikulwat (1998)

EVALUATION OF SIMULATIONS FOR LEARNING OUTCOMES

An analysis of the review of simulation literature identifies learning outcomes instructors adopt as they strive to educate business students. These learning outcomes have been advanced as targeting the skills and knowledge needed by practicing managers. In particular, the sources are (Teach, 1989) (Teach and Giovahi, 1988) (Miles and Randolph, 1985) (Miles et al., 1986) (Anderson, 1982) (Anderson and Lawton, 1997). Simulation researchers have speculated that the method is an effective pedagogy for achieving many of these outcomes. Table 1 below identifies these learning outcomes where

P= measurement of learner perceptions of learning outcomes

O= an objective measurement of learning outcomes

Table 1. Possible learning outcomes for Simulations - adapted from (Anderson and Lawton, 1997)

Facts and concepts of the business discipline				
	Increase the student's knowledge of basic principles and concepts of the discipline			
Interpersonal skills				
<i>Improve the students ability to...</i>				
	Participate effectively in group problem solving	P		
	Motivate coworkers	P		
	Provide meaningful feedback to coworkers	P		
	Resolve conflicts	P		
	Communicate clearly with coworkers	P		
	Develop people	P		
	Lead	P		
	Form coalitions	P		
	Develop consensus	P		
	Delegate responsibility	P		
	Supervise	P		
	Manage People	P		
	Work as a member of a team	P		
	Work in a group environment	P		
	Appraise performance	P		
	Increase the student's knowledge of human behaviour in a group setting	P		
General analytical, critical thinking, problem-solving, or decision-making skills				
<i>Improve the student's ability to...</i>				
	Identify problems	P		
	Frame problems	P		
	Structure unstructured problems	P		
	Analyse problems	P	O	
	Use data to make better decisions	P		
	Distinguish relevant from irrelevant data	P		
	Interpret data	P		
	Implement ideas and plans	P		
	Make decisions using incomplete information	P	O	
	Solve problems	P		
	Solve problems creatively	P		
	Solve problems systematically	P		
	Make good decisions	P	O	
The interrelationships among things				
<i>Improve the student's ability to...</i>				
	Integrate material from various functional areas of business	P		
	See the 'big picture'	P		
	Increase the student's understanding of the complex interrelationships in a business organisation	P		
	Increase the student's understanding of why organisational subsystems must be integrated for organisational effectiveness	P		
Business specific knowledge and skills				
<i>Improve the student's ability to...</i>				
	Assess the situation quickly	P		
	Plan effectively	P		
	Plan business operations	P		
	Schedule and coordinate	P		
	Prioritize tasks	P		
	Forecast		O	
	Use spreadsheet for decision-making	P		
<i>Increase the student's understanding of...</i>				
	The decision process	P		

It is clear from the table above that the vast majority of evaluations have relied on the learner's perceptions of their learning outcomes. Objective measurements (for any learning intervention) are more difficult, however, there appears to be a need to bring in more objective measures to help understand if simulations are an effective method for people to learn business management skills.

CONCLUSION

This paper commenced with three key questions:

1. Why evaluate?
2. What to evaluate? and
3. How to evaluate?

Why evaluate?

Any deliberate learning intervention costs and organisation and individuals' time and money. The value placed on this may vary from person to person, however, there is a value, hence it is worth ensuring that the value of the outcome is greater than or equal to the value of the input.

What to evaluate?

What to evaluate is more difficult to answer. In business and especially management, behavioural competencies as learning outcomes are increasingly recognised as a valid and useful measurement.

How to evaluate?

How to evaluate poses significant problems for the researcher. The training community is most familiar with Kirkpatrick's four levels of evaluation and most frequently measures Level 1 (Reactions) to training intervention as a proxy for all learning outcomes achieved. The basic premise is that if trainees enjoy the training, then they will learn from it. This may be true; however, it does not mean that the trainees have learned what was intended, let alone needed. This suggests that a goal-free (Scriven, 1972) evaluation approach may be suitable to establish what was learned – regardless of what was intended in the intervention.

Greater objectivity about on-the-job behaviour change may be obtained through the use of a suitable instrument and 360° assessment. However, the researcher needs to be aware of Wimer's warning that 360° feedback can have detrimental effects especially when the feedback is particularly negative or not handled sensitively (Wimer, 2002).

EVALUATION STYLE

Choosing an evaluation style requires the researcher to clearly understand his intentions of the evaluation. In order to demonstrate the effectiveness of using simulations as a method of developing managerial competencies, this researcher is interested in 'proving' (rather than learning) the effectiveness of the delivery method and in 'proving' (rather than improving) the development intervention. The experimental research approach to evaluation appears to be the most objective and scientific in approach for this purpose.

RESEARCHER INVOLVEMENT AND BIAS

There are significant problems to consider, those of; the researchers involvement – whether direct or not, the fact of measuring participants in a programme may have a direct impact on the achievement of the learning outcomes. However, if this is the same with all delivery methods being evaluated, the comparison measures on the same basis.

SAMPLE SIZE

The next and very significant issue is that of sample size. It is recognised that in order to achieve statistical validity, the sample size will need to be significantly large – to the order of 100 or more participants.

CONTROL GROUP AND OTHER INFLUENCES

Issues of the suitability of the control group need to be considered as well. It can be expected that individuals will react differently to the same training intervention. Such aspects of the individual as their preferred learning style, their educational history, perhaps gender, perhaps race, perhaps cultural heritage, their age, etc. may all play a part in whether they, as an individual, learn and change behaviour as a result or as a consequence of a particular learning intervention.

The design of a suitable evaluation research method will need to consider all of these issues.

Bibliography

- ALESSI, S. M. (1988) Fidelity in the Design of Instructional Simulations. *Journal of Computer-Based Instruction*, 15, 40-47.
- ALLIGER, G. M., TANNENBAUM, S., BENNET, W., TRAVER, H. & SHOTLAND, A. (1997) A Meta-Analysis of the Relations among Training Criteria. *Personnel Psychology*, 50, 341-358.
- ANDERSON, J. R. (1982) Acquisition of Cognitive Skills. *Psychology Review*, 89, 369-406.
- ANDERSON, P. H. & LAWTON, L. (1997) Demonstrating the Learning Effectiveness of Simulations: Where We are and Where We Need to Go. *Developments in Business Simulation & Experiential Exercises*, 24, 68-73.
- BAILEY, J. & WITMER, B. (1994) Proceedings of Human Factors & Ergonomics Society.
- BEDINGHAM, K. (1997) Proving the Effectiveness of Training. *Industrial and Commercial Training*, 29, 88-91.
- BRINKERHOFF, R. O. (1988) An Integrated Evaluation Model for HRD. *Training & Development*, 42, 66-68.
- BRINKERHOFF, R. O. (1989) *Achieving Results from Training*, San Francisco, Josey-Bass.
- BURGOYNE, J. & COOPER, C. L. (1975) Evaluation Methodology. *Journal of Occupational Psychology*, 48, 53-62.
- BURGOYNE, J. & SINGH, R. (1977) Evaluation of Training and Education: Micro and Macro Perspectives. *Journal of European Industrial Training*, 1, 17-21.
- BUTLER, R. J., MARKULIS, P. M. & STRANG, D. R. (1988) Where Are We? An Analysis of the Methods and Focus of the Research on Simulation Gaming. *Simulation & Games*, 19, 3-26.
- CAMPBELL, D. T. & STANLEY, J. C. (1966) *Experimental and Quasi-Experimental Design for Research*, Chicago, Rand-McNally.
- CAMPBELL, J. P., DUNNETTE, M. D., LAWLER, E. E. & WEICK, K. E. (1970) *Managerial Behaviour, Performance and Effectiveness*, Maidenhead, McGraw-Hill.
- CLARK, R. & CRAIG, T. (1992) Research and Theory on Multi-Media Learning Effects. IN GIARDINA, M. (Ed.) *Interactive Learning Environments; Human Factors and Technical Consideration on Design Issues*. Berlin, Springer-Verlag.
- COLLINS, D. B. (In Press) Performance-Level Evaluation Methods Used in Management Development Studies from 1986-2000. *Human Resource Development Quarterly*.
- DEDE, C. (1997) The Evolution of Constructivist Learning Environments. *Educational Technology*, 52, 54-60.
- DRUCKMAN, D. (1995) The Educational Effectiveness of Interactive Games. IN CROOKALL, D. & ARAI, K. (Eds.) *Simulation and Gaming Across Disciplines and Cultures: ISAGA at a Watershed*. Thousand Oaks, CA., Sage.
- DRUCKMAN, D. & BJORK, A. (1994) *Learning, Remembering, Believing: Enhancing Human Performance*, Washington DC, National Academy Press.
- EASTERBY-SMITH, M. (1980) The Evaluation of Management and Development: an Overview. *Personnel Review*, 10, 28-36.
- EASTERBY-SMITH, M. (1994) *Evaluating Management Development, Training and Education*, Aldershot, Gower.
- EASTERBY-SMITH, M. & ASHTON, D. J. L. (1975) Using Repertory Grid Technique to Evaluate Management Training. *Personnel Review*, 4, 15-21.
- EASTERBY-SMITH, M., THORPE, R. & LOWE, A. (1991) *Management Research: An Introduction*, London, Sage.
- FEINSTEIN, A. H. & CANNON, H. M. (2002) Constructs of Simulation Evaluation. *Simulation and Gaming*, 33, 425,440.
- FILSTEAD, W. J. (1979) Qualitative Methods: A Needed Perspective in Evaluation Research. IN COOK, T. D. & REICHARDT, C. S. (Eds.) *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, Sage.

- GAGNE, R. M. (1984) Learning outcomes and their effects: Useful categories of human performance. *American Psychologist*, 39, 377-385.
- GOPHER, D., WEIL, M. & BAREKET, T. (1994) Transfer of skill from a computer game trainer to flight. *Human Factors*, 36, 387-405.
- GREDLER, M. E. (1996) Educational Games and Simulations: A Technology in search of a (Research) Paradigm. IN JONASSEN, D. H. (Ed.) *Handbook of Research for Educational Communications and Technology*. New York, Simon & Schuster Macmillan.
- GREENO, J., SMITH, D. & MOORE, J. (Eds.) (1993) *Transfer on Trial: Intelligence, cognition and instruction*, Norwood NJ, Ablex.
- GUBA, E. G. & LINCOLN, Y. S. (1989) *Fourth Generation Evaluation*, London, Sage.
- HAMBLIN, A. C. (1974) *Evaluation and Control of Training*, Maidenhead, McGraw Hill.
- HAYS, R. T. & SINGER, M. J. (1989) *Simulation fidelity in training systems design: Bridging the gap between reality and training.*, New York, Springer-Verlag.
- HESSELING, P. (1966) *Strategy of Evaluation Research in the Field of Supervisory and Management Training*, Anssen, Van Gorcum.
- HODGES, M. (1998) Virtual Reality in Training. *Computer Graphics World*.
- HOLTON, E. F., III (1996) The Flawed Four-Level Evaluation Model. *Human Resource Development Quarterly*, 7, 5-21.
- JENKINS, D., SIMMONS, H. & WALKER, R. (1981) Thou Nature are my Goddess. Naturalistic Enquiry in Educational Evaluation. *Cambridge Journal of Education*, 11, 169-89.
- KELNER, S. P. (2001) A Few Thoughts on Executive Competency Convergence. *Center for Quality of Management Journal*, 10, 67-71.
- KIRKPATRICK, D. (1959/60) Techniques for evaluating training programs: Parts 1 to 4. *Journal of the American Society for Training and Development*, November, December, January and February.
- KIRKPATRICK, D. L. (1994) *Evaluating Training Programs: The Four Levels*, San Francisco, Berret-Koehler.
- KIRKPATRICK, D. L. (1998) Evaluating Training Programs: Evidence vs. Proof. IN KIRKPATRICK, D. L. (Ed.) *Another Look at Evaluating Training Programs*. Alexandria, VA, ASTD.
- KRAIGER, K., FORD, J. K. & SALAS, E. (1993) Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78, 311-328.
- MACDONALD-ROSS, M. (1973) Behavioural Objectives: A Critical Review. *Instructional Science*, 2, 1-52.
- MCCLELLAND, D. C. (1973) Testing for Competence Rather Than Intelligence. *American Psychologist*, 28, 1-14.
- MCKENNA, S. (1996) Evaluating IMM: Issues for Researchers. Charles Stuart University.
- MILES, R. H. & RANDOLPH, W. A. (1985) *The Organisation Game: A Simulation*, Glenview, IL, Scott, Foresman and Company.
- MILES, W. G., BIGGS, W. D. & SCHUBERT, J. N. (1986) Students Perceptions of Skill Acquisition Through Cases and a General Management Simulation: A Comparison. *Simulation & Games*, 17, 7-24.
- MOSIER, N. R. (1990) Financial Analysis: The Methods and Their Application to Employee Training. *Human Resource Development Quarterly*, 1, 45-63.
- PATTON, M. Q. (1978) *Utilization-Focussed Evaluation*, Beverly Hills, Sage.
- PEGDEN, C. D., SHANNON, R. E. & SADOWSKI, R. P. (1995) *Introduction to Simulation using SIMAN*, Hightstown, NJ, McGraw-Hill.
- PIERFY, D. A. (1977) Comparative Simulation Game Research: Stumbling Blocks and Steppingstones. *Simulation and Gaming*, 8, 255-68.
- RACKHAM, N. (1973) Recent Thoughts on Evaluation. *Industrial and Commercial Training*, 5, 454-61.

- REDDIN, W. J. (1970) *Managerial Effectiveness*, London, McGraw Hill.
- REEVES, T. (1993) Research Support for Interactive Multimedia: Existing Foundations and New Directions. IN LATCHEM, C., WILLIAMSON, J. & HENDERSONLANCETT, L. (Eds.) *Interactive Multimedia*. London, Kogan Page.
- RICCI, K., SALAS, E. & CANNON-BOWERS, J. A. (1996) Do Computer-based Games Facilitate Knowledge Acquisition and Retention? *Military Psychology*, 8, 295-307.
- ROSE, H. (1995) Assessing Learning in VR: Towards Developing a Paradigm. HITL.
- RUSS-EFT, D. & PRESKILL, H. (2001) *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change*, Cambridge, MA., Perseus Publishing.
- RUSSELL, S. (1999) Evaluating Performance Interventions. *Info-line*.
- SALZMAN, M. C., DEDE, C., R., B. L. & CHEN, J. (1999) A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning. *Presence: Teleoperators and Virtual Environments*, 8, 293-316.
- SCRIVEN, M. (1972) Pros and Cons about Goal-Free Evaluation. *Evaluation Comment*, 3, 1-4.
- SEASHORE, S. E., INDIK, B. P. & GEORGOPOULUS, B. S. (1960) Relationships Among Criteria of Effective Job Performance. *Journal of Applied Psychology*, 44, 195-202.
- SPENCER, L. M. & SPENCER, S. (1993) *Competence at Work: Models for Superior Performance*, New York, John Wiley & Sons.
- STAKE, R. E. (1980) Responsive Evaluation. University of Illinois.
- STANNEY, K., MORRANT, R. & KENNEDY, R. (1998) Human Factor issues in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 7, 327-351.
- TEACH, R., D. & GIOVAHI, G. (Eds.) (1988) *The Role of Experiential Learning and Simulation in Teaching Management Skills*.
- TEACH, R. D. (Ed.) (1989) *Using Forecast Accuracy as a Measure of Success in Business Simulations*.
- THOMAS, R., CAHILL, J. & SANTILLI, L. (1997) Using an Interactive Computer Game to Increase Skill and Self-efficacy Regarding Safer Sex Negotiation: Field Test Results. *Health Education & behavior*, 24, 71-86.
- WARR, P. B., BIRD, M. W. & RACKHAM, N. (1970) *Evaluation of Management Training*, Aldershot, Gower.
- WHITE, B. (1984) Designing Computer Games to Help Physics Students Understand Newton's Laws of Motion. *Cognition and Instruction*, 1, 69-108.
- WHITEHALL, B. & MCDONALD, B. (1993) Improving Learning Persistence of Military Personnel by Enhancing Motivation in a Technical Training Program. *Simulation & Gaming*, 24, 294-313.
- WHITELOCK, D., BIRNA, P. & HOLLAND, S. (1996) Proceedings. IN EDITIONS, C. (Ed.) *European Conference on AI in Education*. Lisbon Portugal, Colibri Editions.
- WIEBE, J. H. & MARTIN, N. J. (1994) The Impact of a Computer-based Adventure Game on Achievement and Attitudes in Geography. *Journal of Computing in Childhood Education*, 5, 61-71.
- WIMER, S. (2002) The Dark Side of 360-degree. *Training & Development*, 37-42.
- WITMER, B. & SINGER, M. J. (1994) Measuring Immersion in Virtual Environments. ARI Technical Report 1014.
- WOLFE, J. (1981) Research on the Learning Effectiveness of Business Simulation Games: A review of the state of the science. *Developments in Business Simulation & Experiential Exercises*, 8, 72.
- WOLFE, J. (1985) The Teaching Effectiveness of Games in Collegiate Business Courses: A 1973-1983 Update. *Simulation & Games*, 16, 251-288.
- WOLFE, J. (Ed.) (1990) *The Guide to Experiential Learning*, New York, Nichols.
- WOLFE, J. & CROOKALL, D. (1998) Developing a Scientific Knowledge of Simulation/Gaming. *Simulation & Gaming*, 29, 7-19.

WOOD, L. E. & STEWART, P. W. (1987) Improvement of Practical Reasoning Skills with a Computer Game. *Journal of Computer-Based Instruction*, 14, 49-53.

YOUNG, M. (2002) Clarifying Competency and Competence. *Henley Working Paper*. Henley, Henley Management College.

John Kenworthy

About the author

John is the Director of Henley Management Education (Asia Pacific) Pte. Ltd., a wholly owned subsidiary of the Henley Management College in Singapore, where he is responsible for marketing and supporting the Henley MBA by Distance Learning in Singapore. He has lived in Singapore since 1997 and considers the island city state his home. John is also Director of Corporate Edge Asia, a company specializing in delivering computer-based simulation and experiential training programmes.

John has a BSC in Hotel and Catering Studies from Manchester Polytechnic and an MBA in Technology Management from the Open University. His research and business interests are in the use of computer-based simulations to develop managerial competence. He is an active member of ABSEL (Association for Business Simulations and Experiential Learning) and AECT (Association for Educational Communications and Technology).

John is a member of DBA 17, commencing his DBA in 2001.