**Evaluating Action Learning: A Critical Realist Complex Network Theory Approach**

Paper bases on a presentation make at the Action Learning Conference, Henley 2010

V2 John G Burgoyne

**Introduction**

This largely theoretical paper argues the case for the usefulness of applying a critical realist, complex network theory (CRCNT) to an understanding of Action Learning (AL) and the challenge of it’s evaluation.

This approach, it is argued, is particularly helpful in the context of improving capability in dealing with wicked problems spread around or located in complex systems and networks (see Boydell and Van Bueren et al. for a discussion of the nature of wicked problems (Van Bueren 2003; Boydell 2007)).

Network theory is the general proposition that the world can be understood as a system of nodes or links at recursive levels (individuals, groups/departments, organisations, clusters and industries etc.) and includes, but is by no means limited to, social networking. Networks can also be events and phenomena, linked causally or otherwise, an approach I will use later here.

Murray Saunders and colleagues (Saunders 2005) argue that evaluation itself can operate as ‘an island of stability in a sea of chaos’. This idea will be developed using this theoretical approach, as will the point that most evaluation is instrumental, about testing the achievement of aims that are assumed to be right and good, rather than testing and challenging these aims – supporting single loop rather than double loop learning, and a non-critical rather than a critical approach in other words. In this part of the argument, and others I will be drawing on Critical Realism (CR) as well at network and complex adaptive systems theory. This is already developed, to some extent based on a presentation to the previous AL conference at Henley in 2008 (Burgoyne 2009).

Amongst other things it is argued that AL can help organisations and groups, understood as networks, balance the destabilising tendencies to explosion and implosion, and, rightly used, can help prevent network distortion (over dominance of one group of stakeholders).

Although largely theoretical, the paper will give an example of an evaluation project on a distributed leadership development initiative with Diabetes Care Networks in the NHS sponsored by the Health Foundation to whom thanks are due. See (Burgoyne 2009).
The structure of the remainder of this paper is as follows:

1. AL, including the issue of what we might expect to be carried forward from AL and hence evaluated.

2. The CRCNT approach.

3. Evaluation of AL from this orientation.

4. The Diabetes Network Development Evaluation as an example, where both the entity being developed is organised on network principles, and these principles were used in our evaluation.

5. Conclusions.

1. **AL, including the issue of what we might expect to be carried forward from AL and hence evaluated.**

It is going to be much easier to consider how to evaluate AL if we have some idea about what we might reasonably expect to be carried forward from it, and therefore know what we are looking for.

I suspect that, in practice, and particularly in corporate settings, AL is often evaluated in terms of success in dealing with the focal project and challenge. This is really treating an AL set as a project group or task force. Important though this is, it is, as I see it, the vehicle rather than the outcome. The aim, as I understand it, is that participants are more successful in future projects and challenges. If there is any truth in the old adage, and I think that there is, that we learn more from our failures than our successes, then this is particularly important. I explored this issue in my previous paper and 2008 conference presentation (Burgoyne 2009).

There are a number of possible lines of thought about what could be carried forward:

a. Firstly is could be the discovery of ‘P’ (programmed knowledge) for what was previously ‘Q’ (questioning knowledge), to use the original Revans distinction (Revans 1998). Q is the stuff of wicked problems, which are this described in contrast to ‘tame’ problems.

This, of course, presumes that some Q, at least, is potentially P, that is troubling problems that are potentially ‘tame’, i.e. open to algorithmic solutions, rather than genuinely and intrinsically ‘wicked’, i.e. not amenable to programmed solutions. It is probably not possible to tell which kind a presenting problem is a priori, the only way to find out is to try.

Even if a solution is found to a problem in a particular situation there is no guarantee that it will travel, or generalise, or transfer either sideways to a different situation or forward over time. According to a CR, complex network theory (CNR and CRCNT) view it may do or it may not, depending on two things: firstly whether the context is
sufficiently similar for whatever the mechanism underlying the potential P to operate in a similar way, and secondly whether the emergent properties of the system work to help or hinder this.

Emergent properties in turn have two possible sources, firstly they could arise from outcomes that are essentially deterministic but arise from combinations of events that have never occurred before, and secondly they could arise from the exercise of agency, or free will, in some form other than the predictable rational optimisation of self interest favoured by classical economists. Related to this is the question of whether P is permanently enduring or has a ‘shelf-life’.

Again, according to the CRCNT view, it could be either. P could be relatively enduring while certain cultural and other conditions persist, and they can for years and decades and more, but when they change the P no longer ‘works’.

Again, it may be difficult to distinguish between absolutely stable background realities and long term cultural and other conditions. Which one the mechanics of capitalism is one of or the fundamental debate between Marxists, in their various forms. Post-structuralism argue that there are no fully enduring structures, and that all are ultimately open to change (Sturrock 1979).

b. Secondly there is the possibility that it is the ability to learn that is enhanced in AL which enhances the capability to deal with future challenges. However what kind of learning? Unless all Q is potential P, how does this work? The Kolb learning cycle (Kolb 1984) seems to presume, implicitly, a rather positivistic view of the world, as the machine the rules of operation of which can be discovered.

We would rather have to see the learning process as a rather subtle kind of detective work inquiring into what kinds of mechanisms could work, how they could be triggered in a particular context and what ‘here and now’ emergent properties might do to affect this. I have briefly developed this idea as one of about sixteen possible schools of thought on learning theory (Burgoyne 2002).

c. Thirdly there is the possibility that AL works through increasing confidence, an argument that also applies to other forms of management education and development. Here, possibly, the realisation that there is a lot of Q in management and that one is not a failure of one does not have the P to deal with every situation.

It is probably reassuring in itself, supported by finding that other people are in the same boat, that being smart and detective like, as in the previous point, is as good as it gets.

d. The fourth and last possibility that I will consider here is that increased knowledge of specific contexts and industries, and the networks involved in them, is itself a form of learning and knowledge that can be productively carried forward.

Research has suggested that this is a benefit from the case study teaching method, often a rival, and a more popular one at least in management education as opposed to management development (see (Pedler 2004) and (Burgoyne 2001).
No doubt there are other possibilities.

2. **The Critical Realist Complex Network Theory (CRCNT) approach.**

Under this heading I will deal with the CR orientation first.

This argues that thinking should start with *ontology* (what exists) rather than *epistemology* (how we can know about it), since the answer to the latter depends on the answer to the former (however to answer the former one is already using some kind of answer to the latter, a real chicken and egg dilemma. I think there is an empirical way out of this that I will explain later).

CR argues that the world is an open system with emergent properties. This in contrast, on the one hand to the positivist position that the world is a deterministic machine or the constructionist (extreme) or post-modern view that world is an ungrounded shifting sea of cultural meaning.

My argument is that if positivism were right then outside the world of physics and engineering, and to a lesser degree biology and medicine, where it has had impressive and world changing results, we would have had more successes with understanding and changing systems that have physical, biological, psychological, social, and (maybe) spiritual elements. Equally if strong constructionism were right then we would have more ability to construct, it think or will into being any world we would like, along the line of *affirmative* post-modernism, as opposed to *nihilistic* post-modernism (nothing means very much or matters very much). Extreme or strong constructionist, post-modernists and post structuralists may argue that there are strong power dynamics in the maintenance and change of meaning and language, drawing on Foucault (Foucault 1980; Foucault 1982; Fillingham 1993); however this implies an underlying reality of power mechanisms which defeats the argument. We do, it seems to me, see events repeat themselves (even regularities in CR speak), but never in exactly the same way, and sometimes very differently. This sounds very much like open systems with emergent properties to me, so I think CR is defensible, at least as the most promising game in town at the moment.

This is highly compatible with the complex adaptive network system view of the world, and we have to be aware of the danger of circularity, i.e. claiming to have proved that we in fact assumed in the position from which we started out. My discussion in the previous paragraph is at least one response to this.

CR has been mainly developed by Ray Bhashar (Bhaskar 1975; Bhaskar 1989) who has gone on to develop the idea in relation to human emancipation (Bhaskar 1986) and in a spiritual and holistic developmental sense (Bhaskar 1986; Bhaskar 1993; Bhaskar 2000; Bhaskar 2002). This theme has been developed in a possibly more accessible way by Archer et al. (Archer 2004). I draw on some of this to develop a more critical and ‘double loop’ approach to evaluation.
Bhaskar in general is not an easy read, and more accessible account of CR are available from Archer (Archer 1998), Sayer (Sayer 1999) and Fleetwood (Fleetwood 2005). It has been applied in the organisation and management studies area by Ackroyd and Fleetwood (Ackroyd 2004) and by myself largely in relation to management, leadership and organisation development (Burgoyne 2000; Burgoyne 2001; Burgoyne 2002; Burgoyne 2008; Burgoyne 2009).

Next, an expansion of this into Critical Realist Complex Network Theory (CRCNT):

Networks are basically made up of nodes and links. These can be people and their relationships, or departments networked to make an organisation, or organisations networked to make up an industry and so on. They can also be events or phenomena with the links being their mutual influences, causal or otherwise.

While one can draw a boundary around a set of nodes and their links, there will always be, arguably, links outside it, hence they are open rather than closed systems (it can be argued that fully closed systems do not exist). Hence most or all systems are open systems with permeable boundaries. In simple systems the casual links are relatively independent as in figure 1:

Figure 1: Relatively Closed Simple Systems:

For example, at the time of my talk at Henley my front door bell stopped working, and my electric toaster stopped working on one side. In the first case the battery had gone flat and in the second one of the grills had fused. These were independent events, solvable separately, by replacing the battery and the grill respectively. Solving one had no effect on the latter.
But in complex systems the relationships are much more entwined, as in figure 2:

Figure 2: Relatively Complex Open Systems:

Duncan Watts, in his book ‘Six Degrees: The Science of a Connected Age’ (Watts 2003) one of the more readable books on network theory, describes a breakdown of a large part of the electricity supply system in the United States. This was triggered by a tree branch touching some overhead supply lines on a relatively minor link in a period of high demand. Normally this would not have happened, but a set of circumstances in which a number of backup systems were out of operations, and others vulnerable, meant that it escalated, like a virus epidemic spreading, round the system and covering a whole state and more (another example).

It is as though my front doorbell battery running flat brings down a whole set of systems across the Lake District in Cumbria, England, where I live.

3. Evaluation of AL from this orientation.

There are two main challenges for evaluation, other than deciding what is a ‘good’ purpose to evaluate against, to be discussed later. The first is the measurement of change and the second is the attribution of causation. Both are challenging, but, arguably, the latter is the more so. Figure 2 helps explain why. If the intervention is one of the causes at the bottom and the desired outcome is one of the effects at the top, then there is more than one route from the
former to the latter, other causes at the bottom may influence the effect in question at the top and emergent properties in between may also influence the outcome.

Returning to the question of evaluating AL, and building on this, what we could look for in evaluation interviews with AL participants is the ability to read complex adaptive network systems at two levels. Firstly in terms of the specific situations they are dealing with, and secondly their ability to generalise this to other situations.

What we could look for in evaluation interviews with AL participants is the ability to read situations as complex adaptive network systems.

This would be an example of starting with the clinical and moving to the epidemiological. After collecting a number of such accounts one could look for any common reported effects, as seen through self observation and self report, and seeing if these suggest any more objective and external measures that could be investigated more quantitatively, either ‘soft qualitative’, as in perceptions of team quality in the diabetes study, quantified on rating scales, or ‘hard quantitative’ as exemplified by the percentages of detected diabetics.

In this case networks are events and causal links, with an eye to intended outcomes and events open to intervention. In putting a boundary around part of this causal event network one can also look out for variations with things beyond this boundary, as in the England Scotland/Wales contrast in restructuring and reorganisation.

Another approach to evaluation, informed by this general approach, is to plot the intended process and outcome journey of the designers of the intervention to be evaluated, in terms of both processes and effects, this I will call an ‘intended event chain’, as illustrated in figure 3, which comes from another area entirely, a study of the effects of customer care initiatives on organisational performance:
Figure 3: an intended event chain:

An example of an event chain:

From an evaluation of customer service interventions:

Evaluation can then proceed by studying how the ‘actual’ compares with the ‘intended’ by whatever means seem suitable.

Another example comes from the National College for School Leadership (now just the National College) set up to develop leadership in all parts of primary and secondary education.

They developed the model in figure 4 as a general roadmap of how they hoped that the College would work, in general through its whole portfolio of programmes and initiatives.
Figure 4: Evaluation Framework for the National College for School Leadership:

**National College for School Leadership**

1. *Reach* – the extent to which the College offers leadership development support mechanisms to its constituency.

2. *Engagement* – the extent to which the constituency engages with the mechanisms on offer.

3. *Learning* – the extent to which learning occurs relevant to leadership practices.

4. *Application* – the extent to which helpful changes in leadership practice in the target domains takes place.

5. *School improvement* – the extent to which improvement in performance of schools, or part of them, can be detected and reasonably attributed to improved leadership practice.

6. *System development* – the extent to which improvements in the education system, of which schools are components, can be detected and similarly attributed to improvements in leadership practice.

These six stages are intended to be sequential, a chain of events that may or may not happen, or get so far and then be blocked. Again, this serves as a framework for determining what to look for in terms of measures, and what processes to inspect for to establish causality.

This framework is probably appropriate for a start up initiative, and one where there is a degree of voluntarism about its use (though this became more limited, as, for example, qualification became mandatory for newly appointed school heads, a contextual factor to look out for in itself).

This example can be seen as a development of the Kirkpatrick (Kirkpatrick 1983; Kirkpatrick 1996; Tamkin 2002) model, arguably the most known and used on in the whole evaluation field. It is shown in figure 5.
This, again, is a causal chain, like a greasy pole, that an initiative can crawl along and get to the end or fall off at one of the intermediate stages.

Figure 6, below, suggests a four level stairway to best practice in evaluation. I find it helpful to distinguish small ‘e’ evaluation, natural and informal that always goes on, from big ‘E’ Evaluation that is deliberately and systematically planned and carried out. This is about Evaluation.
Many people think that level three is the top of the stairway. The suggestion of the fourth level comes from Ray Pawson (Pawson 2002). Pawson, along with Tilley, have developed a CR approach to evaluation (Pawson 1997), which, incidentally, provides a relatively rare example of what CR looks like in practice, much of the CR literature, including that cited above, remains at the level of meta-theoretical and philosophical discussion.

Pawson’s argument is, in effect, for evaluation led design, that is to say that design of any new initiative should start with a review of what is already known, and preferably research and evidence based, that is relevant to that design.

Pawson, working out of a CR orientation as is the case with this paper, recognises that this is not just a straightforward matter of generalising forward from things that have worked in other contexts in the pasts, which would be the positivist approach. In the positivist approach this is usual called ‘meta-analysis’ where data from similarly designed studies, such as clinical trials on a medical intervention, or correlations of a psychological test with some kind of outcome measure are combined so that they are, in effect, one large sample instead of several little ones. Nor is it an entirely qualitative or constructionist one, which is, in effect, a literature review.

Pawson proposes ‘realist syntheses as the preferred approach, concentrating of research that is about the same presumed mechanism, irrespective of the particular area of application. One example he uses is surveillance as a means of control (which Foucault is quite critical of as a general social process (Foucault 1977). Here it may be surveillance cameras in car parks or shopping centres, ‘naming and shaming’ initiatives, publication of performance league tables and so on.
There is usually no shortage of literature on any particular topic. Only a small proportion of this, typically 10-20% at best, is empirical and evidence based, and of this small proportion much of it is questionable in terms of its design and rigour.

Figure 7 shows my suggestion of levels from less useful to more useful, to grade such studies in a realist synthesis literature review:

**Figure 7: Levels of Usefulness in a Realist Synthesis Literature Review**

*Grading literature about this causal chain accounting*

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Quantitative correlation with evidence of causality from left to right</td>
</tr>
<tr>
<td>2.</td>
<td>Quantitative correlation without evidence of causality from left to right</td>
</tr>
<tr>
<td>3.</td>
<td>Strong qualitative data of influence from left to right</td>
</tr>
<tr>
<td>4.</td>
<td>Weak qualitative data of influence from left to right</td>
</tr>
<tr>
<td>5.</td>
<td>Strong logical or theoretical argument of influence from left to right</td>
</tr>
<tr>
<td>6.</td>
<td>Weak logical or theoretical argument of influence from left to right</td>
</tr>
<tr>
<td>7.</td>
<td>Unsupported personal opinion or polemic about the above</td>
</tr>
<tr>
<td>8.</td>
<td>Not relevant</td>
</tr>
</tbody>
</table>

As used here, ‘left’ is like the bottom of figure 2, to do with interventions and causes, and right is to do with effects and outcomes, as at the top of figure 2. It is ‘left’ and ‘right’ as set out in figure 3 to do with customer care initiatives.

In a realist synthesis some kind of weighting would be given to a study based on its ranking in these terms.

**What else from complex adaptive network system theory?**

a. Thinking about what it takes to move to more sustainable and more desirable equilibrium state. There is a potential paradox here. Something that is easy to change is also more likely to change back again through natural processes. As a Trustee of Brathay Trust, an outdoor management development charity, specialising amongst other things in troubled and troublesome youth, I observe that it is possible to change their attitudes and intended behaviours, but once they are back in their inner city gang environment this soon changes back again. This is why these programmes now include mentoring and the like back in these situations, and other attempts to change
this context which, while not being the whole answer by any means, gives the programmes overall a better chance.

b. Another answer, from complexity theory, is the idea of moving a system to far from equilibrium positions. If you move a stone a little way up the side of a valley it is likely to roll down to the bottom again when you leave it. If however you move it to, and over, the top of the hill between two valleys it will roll to the bottom of the next one. This, of course, presumes that the ‘grass really is greener’ in the next valley, and you may not know till you get there. The idea may not always be easy to apply, but it does at least suggest an approach.

c. There may be situation where the stone is already near the top of the hill, and it may not take much to nudge them over (or back), depending on circumstances. These are what are often referred to as ‘tipping points’. Again, a useful idea that may not always be easy to use.

And again, there is the question of whether we are morally right to do this in any particular situation. By what right do we think we know best? And what about other actors in the situation that may be pushing for different outcomes by the same or different means?

**Critical Evaluation?**

This takes us on to the observation that most evaluation is instrumental, i.e. does it work against taken and unquestioned objectives and purposes?

What about evaluating whether we are doing better things as well as doing things better?

And how do we know what is better? Some philosophers (Hulme for example(Hulme 1739)) argue that one cannot get from ‘is’ to ‘ought’ by research and science. Science can find out if A can lead to B, but cannot establish whether B is a good thing or not.

I do not agree, though I have to take a position and make some assumptions to do so. Mine is a combination of Darwin’s classic evolutionary view (Darwin 2003), as modified by Maturana and Valera’s (Maturana 1980; Maturana 1992) idea of autopoiesis, Bateson’s modification, also to Darwin, that progress is not just random variation, natural selection and reproduction but moving to larger units of survival (Bateson 1972), and Lovelocks proposition (the Gaia Thesis) that the whole universe is a living, evolving (and learning) system (Lovelock 2000).

In terms of Darwin vs. Maturana and Valera, and the debate as to whether the focal case adapts to the environment or vice versa, I propose a synthesis in which there is a process of co-evolution, thus elements of both. Though as Binns (Binns 1994) the more ‘advanced’ the entity, e.g. human being rather than amoeba, the more it is likely to be autopoetic. This thought is also reflected in Senge’s (Senge 1990) suggestion of a progression, in terms of organisations’ abilities to learn, from the **reactive** to the **adaptive** to the **generative**. In these terms the adaptive is Darwinian, and where much of the strategy literature is, and generative is autopoietic.
Being practical, evaluators ought at least to question the ‘given’ purposes and objectives against this or other criteria. It should ask what stakeholders in the situation they are working for and who the other stakeholders are, and what their interests are, and how they see the legitimacy of all these.

Often evaluators may not be able to resolve all these issues, but, as a variant on Saunber’s idea as evaluation as an ‘island of stability in a sea of chaos and complexity’ (Saunders 2005), and my ‘arena thesis’ (management development as a arena where different values and interpretations can be negotiated), an evaluation project may be a place where different stakeholders can come together to try to sort things out. This may not always be easy or possible, but it is at least worth a try, and talk and compromise, even if only temporary, may be better than conflict and war (literally and metaphorically), as argued by Isaiah Berlin under the heading of ‘agonistic liberalism’ (Berlin 1969).

**The Diabetes Network Development Initiative Example**

The Diabetes example, as a case study in evaluation will be discussed in the next section, but, as the initiative was to evaluate an intervention to influence a network organisation, it will be used here as a substantive example. In terms of network theory, networks can ‘implode’, which is what happens when the internal links are stronger than the external ones, at the extreme this is a cult where virtually no external links are allowed.

‘Explosion is the opposite, and occurs when the external links are stronger than the internal ones. At the extreme, here the networks disappear altogether.

My observation of the diabetes networks was that they were more likely to explode than implode, medical clinicians being more loyal to the British Medical Council, nurses similarly to their professional body, and administrators to Trust Chief Executives and so on.

An advantage of AL is that it may counterbalance the tendency for networks to go one way or the other. This by both bonding the network members but also strengthening their links to the external stakeholders in the challenge that the AL set is addressing.

Finally, though there is more of a value judgement here about what is ‘normal’ or ‘right’ or ‘desirable’. Here, interestingly, medial networks, despite various initiatives about patient participation, consultation and involvement, my reading is that they are still distorted away from the patients towards the professional bodies and the administrative hierarchy.

4. **The Diabetes Network Development Evaluation as an example, where both the entity being developed is organised on network principles and these principles were used in our evaluation.**

In our evaluation study of interventions to help diabetes networks in the NHS (Burgoyne 2009) we borrowed an approach from the more positivist research method portfolio, which we called ‘comparator’ groups (control groups is the normal positivist term). Our diabetes
networks covered geographical areas, one of which was Southport on the UK Lancashire coast, a middle class retirement and holiday resort. We chase Formby, also on the Lancashire coast and with similar demographics as the comparator. Similarly for large urban sites, rural sites, and poorer inner London Boroughs.

Interestingly, in terms of the effect that did show a difference, the diabetes teams’ perceptions of the quality of their teams, the intervention groups stayed the same while the comparator groups went backwards, for the intervention groups studied in England. In Scotland and Wales it was the other way round. This led us to ask why and whether there was a systematic difference in context. An obvious candidate was the amount or restructuring and re-organisation going on around the diabetes network teams. This was much greater in England than in Scotland and Wales. This was the quantitative, epidemiological finding (looking at the correlation between the inputs and outputs to the ‘black box’. To check this we looked clinically and more qualitatively at the situation, equivalent to opening the lid of the black box and seeing what, processually, at what is going on. In one of the English diabetes networks we found that due to restructuring the number of medical clinical posts (doctors as opposed to nurses etc.), was reduced from five to three, and the five incumbents has to apply for the three posts. Hardly conducive to good team relationships as their stories showed.

This is an example of mixing the quantitative, epidemiological, and the clinical, qualitative. In terms of figure 2, looking at correlations between cause and effect variables, and then for processual stories that illustrate the connections and pathways between.

Another example from the diabetes studies: we found no effect in terms of hard data, for example the percentage of diabetics detected by the networks as a percentage of the number predicted demographically to be there. We wondered why: no effect, measured too early, different from the aims of the networks, insensitivity of the measure. We found some qualitative data to support the last of these. One of the intervention groups got organised, with their OD consultant’s help to put on a ‘road show’ e.g. hiring a village hall and inviting passersby in for a diabetes check-up. This lead to a few dozen previously undetected diabetics being found. However this would not show up on the overall data used for a whole regions where the numbers were thousand or tens of thousands. Notice that this ‘qualitative’, clinical form or research does not literally mean not using numbers, as in this case, it is what they are used for and how they are used that makes the difference.

In terms of the use of comparator and control groups there is little difference between what one does in terms of method from the positivist or CR point of view. The main difference is in generalising the conclusions and the attention paid to context. CR would be much more cautious about generalising from what has happened in one situation to others elsewhere and into the future.

To end with the findings from our diabetes study, we found an effect, in terms of a differential between our intervention sites and our comparator sites, in terms of the soft quantitative data, i.e. members’ perception of team quality. We did not find an effect at the level of hard quantitative data, percentage of diabetics detected, diagnosed and treated in
various ways, but this may be due, amongst other things to the insensitivity of the data used, as discussed previously.

In terms of how the intervention had an effect, it made no difference which of the usual tools the consultants used: Myers-Briggs, Belbin, Tuckman etc. What mattered was very basic: whatever got them to know each other a bit better, to bother with meetings, to have agendas for them, to have them properly chaired, and to have minutes with action points that were followed up.

5. In Conclusion

I hope I have convinced you that critical realist complex network theory provides a useful basis for evaluating AR, amongst other development initiatives.

This is for two main reasons, in no particular order of importance:

Firstly it offers a way of dealing with causality, when there are many other sources of influence other than the initiative being evaluated, more than one route between cause (or intervention) and effect (or outcome), emergent properties on the journeys between them, and variations according to context.

Secondly it allows an integrating mix of the qualitative and quantitative, the clinical and the epidemiological in medical research terms, which can usefully be applied iteratively with each giving clues as to what to look for with the other approach, thus helping to structure a programme of evaluation research rather than just the individual projects that make this up.

The stairway model in Figure 6 suggests that such a programme can start with a ‘realist synthesis’ of previous evaluation research, in order to use this to the maximum benefit to feed into the design of the intervention. Such a review is based on CR principles in contrast to the meta-analysis of a more positivist approach and the qualitative literature review characteristic of the more constructionist approach. It is generally thought that such a review should concentrate on studies where there is the same presumed change mechanism at the heart of it, irrespective of the intended outcome and context, though this can be contested. Either way it should end with the initiative being planned on the basis of existing knowledge and the most appropriate ‘before’ and ‘process’ measures identified for the evaluation.

I hope that all this is both philosophically satisfying but practical and doable.

My practical proposition, following on from this, is to start by interviewing participants along the lines I have suggested, and work out from there to any harder and more objective measures that this might suggest.

I have also proposed that evaluation should add a critical strand, providing the forum, or arena, for a discussion amongst various stakeholders about aims and purposes.
References


Saunders, M., Charlier, B., and Bonamy, J. (2005). "Using evaluation to create 'provisional stabilities':
bridging innovation in Higher Education change processes." Evaluation: the International Journal of
Theory, Research and Practice 11(1).


Doubleday.

University Press.

Brighton, Institute of Employment Studies.

Problems in Networks: Analyzing an Environmental Debate from a Network Perspective." Journal of
Public Administration Research and Theory 13.