

Six Normative Approaches to Evaluation

Jess Dart

What is ‘Good’ Program Evaluation?

Evaluation is used in many program contexts, in many countries and across many different disciplines. It is also used for many different purposes, and the range of things that are included under the umbrella of program evaluation is considerable. Even within one project there may be several evaluation initiatives under way. For this reason, in ‘good’ evaluation the choice of evaluation approach needs to be context-specific and take into consideration the purposes for which the study is being undertaken. This view is endorsed by many evaluators including Worthen et al. (1997) and Owen (1993).

Notwithstanding such recognition, several authors have advanced normative models for what they believe entails good evaluation. Predictably, there are considerable differences between these models. Despite the differences, an examination of these models can offer many insights into what entails good evaluation. Therefore, present six different approaches to evaluation and from these develop a series of premises for good evaluation. I propose a classification of evaluation consisting of the following six normative approaches, each of which will be discussed in turn:

- i. Experimental approaches
- ii. Testing-objectives approaches
- iii. Decision-management approaches
- iv. Judgemental approaches
- v. Pluralist-intuitionist approaches
- vi. Theory-driven approaches

These six approaches are an amalgam of the meta-models of Smith (1994), Stake (1973), House, (1978), Stufflebeam and Webster, (1981) and Worthen et al. (1997). In reality this classification, and the ones it is based on, are over-simplistic because of the complexity of the models they attempt to classify. Wadsworth (1991: 62) points out that ‘One would really need a three-dimensional map on which to try to plot a full picture and even then some of the techniques would have to be moved back and forth between

one category and another'. However, the meta-model proposed does allow an examination of the range of things that theorists contend are important for 'good' evaluation.

The theories behind both of the first two approaches in the meta-model (experimental approaches and testing-objectives) have been partly superseded by more modern approaches. However, they are included in the meta-model because they are still seen in practice and because critique of these approaches paved the way for the more contemporary approaches; theory-driven, judgemental, decision-management and pluralist-intuitionist.

i. Experimental approaches

Experimentalism, often referred to as the 'classical paradigm' in program evaluation, is characterised by the work of Popper (1959), Campbell, (1991), and Cook (1966). Experimental approaches construe evaluation as a knowledge generation research task. Therefore good evaluation creates knowledge and theory about a program situation that can be ratified by empirical data. Campbell focused much of this energy on the notion of minimising the 'threats to internal validity', and developing the most plausible explanations for the results of experiments.

Experimental approaches are based on a Humean theory of causation. The basic task is to hypothesise or demonstrate the constant conjunction whereby action 'X' produces outcome 'Y'. Advocates of such approaches promote the use of experimental and strong quasi-experimental designs that provide the strongest causal inference. Experimental designs in evaluation in principle follow the same rules as agronomic trials: the 'treatments' are randomly assigned (in this case to program participants) to either a 'treatment' group or to a control group. Experimental approaches may also have pre- and post-test designs, so that changes can be monitored before and after the treatment period.

Experimental designs offer a logical approach for determining whether certain program variables affect program outcomes. While experimental approaches can be criticised on several fronts, there is little doubt that experimental evaluation paved the way for important insights into evaluation theory, for example, the work of Campbell (1991, 1969) and Cook (Cook and Campbell, 1979) who focused on the 'internal threats to validity' associated with experimental design.

However, for several reasons, experimental designs did not always prove their value, and many

expensive designs yielded inconclusive findings. Returning to the analogy of agronomic trials, the secret to a good experiment is to ensure that the plots are randomly allocated and that all other variables are controlled. Even with plants, this can produce results that are difficult to apply to the reality of the farmers' fields, as the controlled conditions of the research station do not correspond to real life conditions. It can be even more difficult to control variables amongst human participants and it is often unethical to subject one group of people to a treatment, and deny it to others. Even where large trials have been conducted, the results have been disappointing. Variables between the responses of different people within the same treatment group are seen as 'noise' in the experiment. Therefore, the results of an experimental design only provide information as to whether the program 'worked' for the entire group of people or not. It provides no clues as to why the program may work for some individuals and not for others.

Indeed, for practical purposes, experimental designs often exclude many of the contextual factors that influence cause-and-effect relationships. It can be argued that these contextual factors are the very thing in which evaluators should be most interested. Theory-driven approaches to evaluation for example, reject a key notion inherent in the logic of experimental evaluation in that it cannot fully take into account either the key mechanisms linking programs with outcomes or the richness of heterogeneous contexts (Feinstein, 1998). Indeed, theory-driven evaluators such as Pawson and Tilley (1997) attack experimental evaluation for yielding very little in terms of learning about programs. Their reasoning is:

By its very logic, experimental evaluation either ignores underlying process, or treats them incorrectly as inputs, outputs or confounding variables, or deals with them in a post hoc and thus arbitrary fashion (Pawson and Tilley, 1997:54).

Despite their limitations, experimental designs are still used and valued in many circles – especially in the United States.

ii. Testing-objectives approaches

Evaluation approaches in the testing-objectives category are focused on determining whether the stated goals or objectives of a program have been achieved. Tyler (1967) was among the first to develop this approach which he referred to as 'educational evaluation'. Good evaluation under this model depends on being able to accurately determine the extent to which stipulated objectives have been reached. Tyler defined evaluation as the process of determining the extent to which the objectives of a program are attained (Worthen 1997:82).

Guba and Lincoln (1989) relate, how in 1933, Tyler was engaged to carry out an eight-year study of an alternative school curriculum. The idea was to collect information about the extent of pupil achievement against defined objectives, to guide revisions to the curriculum. The results of each trial were not available until after the trial was complete. This process was iterated over successive course offerings until the curriculum was found to produce an appropriate level of achievement. In a sense, this was formative evaluation and the evaluators' role was that of describing the attainment of objectives over time. A point to note about this approach, however, was that the results of the evaluation were not made (publicly) available until after the program was complete. This implied that the program could not be modified during implementation. This approach to evaluation is still used in extension evaluation nowadays.

Objectives-testing approaches differ from the previously popular model of 'comparative experiment' in that they do not involve the expensive and disruptive comparisons between experimental treatments and the control. Since Tyler's approach calls for the measurement of behaviourally defined objectives, it concentrates on learning outcomes instead of organisational and teaching inputs (Madaus et al., 1983). Educational evaluation that predated objectives-testing approaches tended to focus on measuring the performance of pupils, rather than examining the curriculum or program itself.

Careful articulation of the objectives of a program – an essential part of an objectives-testing approach – can have positive ramifications for both program planning and delivery. Very few evaluations are conducted without some consideration of the extent to which objectives have been achieved.

However, critics of objectives-testing approaches found it unsatisfactory to be unable to make corrections to the program until it was complete. In other cases, practitioners were unwilling to stipulate pre-determined 'outcomes', as they were unsure at the onset of the programs about the appropriate outcomes. Both of these criticisms can still be heard today of evaluations that are strongly outcome (or objectives) focused. Another important criticism of this model was levelled by Stake (1967), who pointed out the neglect of judgement in the objectives-testing model. He suggested that evaluation requires standards against which judgements can be made and that the inclusion of standards must be value-laden. Scriven (1967) also pointed out that the objectives themselves should also be treated as problematic and subject to scrutiny. His argument was that testing the extent to which the goals of a program have been achieved does not determine the worth of that program for society in general; it does not assess the goals themselves.

iii. Decision-management approaches

Decision-management approaches aim to serve decision-makers' needs in managing programs. Evaluation models fitting into this category are Patton's Utilization Focused Evaluation (1997), and Stufflebeam's CIPP model (context, input, process, and product) (Stufflebeam and Webster, 1981), that relies on a modified systems analysis approach. Good evaluation, under this approach, would produce findings that are used to bring about effective decision-making for the organisation. This type of evaluation is largely judged by how well it has served those commissioning the evaluation. For example, (Caron, 1993: 62) states that evaluation is 'carried out for the organisation. It is a function of management'. Therefore, in evaluation of this genre the aim is to serve the needs of program managers rather than of the program's clients or the wider public.

Patton, who is probably the most well cited advocate of decision-management approaches, defines program evaluation as:

...the systematic collection of information about the activities, characteristics, and outcomes of the programs to make judgments about the program, improve the program effectiveness, and/or inform decisions about future programming. Utilization focused program evaluation (as opposed to program evaluation in general) is evaluation done for and with specific, intended primary users for specific, intended uses (Patton 1997: 23).

According to Patton (2000:1) Utilization Focused Evaluation is an 'approach to making evaluations useful, practical, accurate, systematic, and ethical'. This involves matching the evaluation approach and the design to the information and decision needs of primary intended users, taking into account 'other stakeholders, political factors, organisational constraints, project/program history, available resources, and cultural factors of a specific evaluation context' (ibid: 1). He suggests that this allows for 'situationally responsive' evaluations.

While decision-management approaches offer sensible suggestions for increasing the likelihood of the evaluation findings being used, they have been criticised for the apparent 'cosy' relationship between the evaluator and the project management. For example Pawson and Tilley (1997) caution that these pragmatic approaches 'suggest a Rothschildian vision with research skills for hire' – the Rothschild principle being:

the customer says what he wants, the contractor does it if he can, and the customer pays (Department of Health, 1993; cited by Pawson and Tilley 1997: 14).

Thus, decision-management models are challenged on their ability (or inability) to present unpalatable information to the management. Some decision-management models can also be questioned on their assumption that evaluation serves the needs of program managers and staff rather than program clients or the greater public.

iv. Judgement approaches

Judgement approaches involve the professional judgement of experts. This approach includes Scriven's (1976) 'Goal-free' evaluation and Eisner's (1985) Connoisseurial model of evaluation. Here, evaluation is seen as a determination of the merit or worth of a program – and an evaluation of this type would be judged on the basis of the accuracy and lack of bias in conducting a judgement of worth or merit of the program. Definitions of evaluation for this are characterised by the words of Scriven:

It's extremely important not to over simplify the logic of evaluation by defining evaluation as, for example, the provision of information to decision-makers. Evaluation is what it is, the determination of merit or worth, and what it is used for is another matter (Scriven, 1980: 8).

Bad is bad, and good is good, and it is the job of evaluators to decide which is which. (Scriven, 1986: 19)

An extreme model from this genre is Scriven's 'Goal-free Evaluation'. 'Goal-free evaluators begin evaluations totally blind' (Scriven, 1976: 137) to stated goals. Evaluators have to discover what effect the program has and match their effects against the needs of those who they affect (Scriven, 1976: 137). The 'Goal-free' evaluator avoids contact with program staff, who may bias the conceptualisation of the evaluation questions. The evaluator is required to judge the merit of the program in terms of empirical evidence of the effects of the program intervention.

In Goal-free Evaluation, Scriven attempts to address the problem of over-emphasis on program outcomes (or objectives), stating that they should be totally ignored. The evaluator's job, according to Scriven, is to locate any and all program effects, intended or not, that might help solve social programs. His argument is that testing the extent to which the goals of a program have been achieved does not determine the worth of that program for society in general; it does not assess the goals themselves. Goals are a poor source of such effects, and are 'often vaguely worded to muster political support and rarely reflect side effects that are difficult to predict' (Shadish et al., 1995:80).

Goal-free Evaluation does serve to remind evaluators of three very important points. Firstly, in addition to examining stipulated outcomes, evaluation should look beyond the goals of the program itself, ie., it

should consider unexpected outcomes. Secondly, in some situations it may also be necessary to question the value and logic of the outcomes themselves. A third contribution of Goal-free Evaluation is the concept that good evaluation involves some judgement of merit or worth of the program's impact.

However, Shadish et al. (1995: 114) suggest that 'some evaluators have difficulty accepting the notion that they can, much less should, evaluate a program without knowing its goals'. While most evaluators have heard of Goal-free Evaluation, they may not see it as central to their thinking about evaluation, and they still use goals as the most common source of dependent variables (Shadish and Epstein, 1987). Another criticism of the Goal-free Evaluation is that while it may be a very useful theory, it is not necessarily a practical model. Indeed, few cases of Goal-free Evaluation have been documented.

Critics of Goal-free Evaluation consider the term 'goal-free' evaluation to be a misnomer. The evaluator does not get rid of all goals, but replaces the goals of the project staff with more global goals based on societal needs and basic standards of morality (Alkin, 1972 cited by Patton 1997:183). Patton argues that Goal-free Evaluation:

...eliminates only one group from the game, local project staff. He (*Scriven*) directs data in only one clear direction – away from the stated concerns of the people who run the program. He addresses an external audience such as legislative funders. But, in as much as these audiences are ill defined and lack organisation, I am unconvinced that the standards he applies are none other than his very own preferences about what program effects are appropriate and morally defensible. Goal-free Evaluation carries the danger of substituting the evaluator's goals for those of the project (Patton 1997:182).

v. Pluralist-intuitionist approaches

Pluralist-intuitionist approaches share a common commitment to value pluralism; that is to identify and preserve multiple value perspectives. This includes the models of Stake (1967), Guba and Lincoln, (1989) and Wadsworth, (1991) and the participatory monitoring and evaluation approaches to evaluation that are seen in development projects (Estrella and Gaventa, 1998). All the approaches in this category are highly client centred and present a subjectivist ethic and epistemology within a liberal ideology (Smith, 1994). Good evaluation under this normative model involves creating increasingly sophisticated and shared constructions of reality. For example, evaluators employing Guba and Lincoln's 'Fourth Generation Evaluation' (1989) judge evaluation by the extent to which it succeeds in involving a range of participants in sharing their views. Guba and Lincoln, (1989) for example, advocate that evaluators ought to strive for, not validity, but increasingly sophisticated constructions of reality. Some versions of

pluralist-intuitionist approaches also focus explicitly on social justice, and liberation of program beneficiaries from oppression.

Pluralist-intuitionist approaches evolved as a reaction to the limitations of the logical positivist approach to knowledge generation such as that adopted in experimental approaches. To overcome the limitations of these approaches, a number of overlapping frameworks emerged that were based on the tradition of anthropology. Pluralist-intuitionist approaches include interpretivism, relativism, naturalistic inquiry, constructivism, and feminist inquiry. What they hold in common is the rejection of the existence of a singular knowable reality. This ontological position has significant ramifications for the field of evaluation; evaluation is concerned with ‘apprehending reality’, thus questioning the nature of ‘*reality*’ is clearly going to affect evaluation practice at several levels. In Chapter 5, two models of this genre are examined in more detail.

One family of approaches that fits into this genre is the Participatory Monitoring and Evaluation approaches (PM&E), also known as Participatory Learning and Action approaches. These approaches are used increasingly in development projects. This type of evaluation is not generally referred to in the main texts on program evaluation, but is gaining increasing attention in the international literature of agricultural extension (Estrella and Gaventa, 1998; Guijt and Gaventa, 1998). Like Fourth Generation Evaluation, PM&E approaches share a common commitment to participation of the beneficiaries of a program. However, PM&E approaches tend to emphasise action a little more strongly, and to be influenced by models of action research. Good evaluation under PM&E approaches would be based on the worldview of the people it is aiming to serve, and would encourage participants to take self-directed action to improve their own or collective social conditions. PM&E approaches were influenced by the Brazilian activist Friere, (1972) and the work of action researchers such as Lewin (1948) and Whyte (1991).

PM&E approaches were developed in reaction to dissatisfaction with the rapid, yet participatory approaches named ‘participatory rural appraisal’ (PRA), and the limitations and dangers of reliance of non-participatory forms of monitoring. PM&E approaches are largely qualitative, participatory approaches with a focus on organisational learning and empowering the beneficiaries of the project.

Pluralist-intuitionist models reinforce the notion that in good evaluation, a range of different perspectives should be included. The evaluator is reminded that programs are conducted in a political environment, and the values of the different stakeholders must be brought to the surface, not excluded or ignored.

Pluralist-intuitionist approaches offer a way of appreciating and incorporating the different ‘theories of action’ or ‘constructions’ of the project held by different stakeholders. These approaches can be especially valuable in the sort of evaluation that involves a broad range of stakeholders involved with disparate views. They are often highly useful in participatory programs, and particularly when the program intervention emphasises empowerment.

Like other approaches, pluralist-intuitionist models are not appropriate to every evaluation context. In some cases, evaluations of this type may not meet the client’s information needs. Two examples illustrate this point. Firstly, the client of the evaluation may not entertain the view that there are many possible versions of ‘reality’ but instead may request a definitive account of the ‘facts’ about program success. Secondly, clients who require firm recommendations to be a part of the evaluation output may be dissatisfied with this approach. Evaluators who conduct pluralist-intuitionist evaluation generally avoid making recommendations, and the evaluator may refrain from offering concrete ‘recommendations’, and see their role as more of a ‘facilitator’ than a ‘judge’.

Many of the pluralist-intuitionist approaches, such as Fourth Generation Evaluation, strive to gain consensus between all stakeholders. This can be questioned on two fronts. Firstly, the achievement of consensus is not always a realistic goal, especially when different stakeholder groups may have strong political differences. Secondly, the achievement of consensus is not always a helpful goal, especially if it is won at the expense of ‘less powerful’ or minority opinions being silenced in favour of the majority opinion.

vi. Theory-guided approaches

Recently there has been a growing interest in theory-guided approaches to evaluation (Pawson and Tilley, 1997; Chen, 1990; and House, 1991). These approaches involve the construction of ‘program theory models’ – casual models that elaborate how a program is intended to achieve intended outcomes. Bickman, (1987: 5) defines program theory as ‘the construction of a plausible and sensible model of how a program is supposed to work’. Chen explains that program theory has both descriptive and prescriptive concerns and defines program theory as:

A specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated (Chen 1990:43).

Good theory-driven evaluation would, therefore, develop highly plausible theory concerning how a

program works, in what situations and why. This theory would be the best explanation of empirical data, and any alternative contending explanations would have been refuted by the empirical data. In Chapter 5, one model of this genre is examined in more detail: Pawson and Tilley's (1997) Realistic Evaluation.

The advantages of theory-driven evaluation are that it goes beyond establishing whether a program works, allowing an understanding of how a program works. Wortman (1983: 20) comments that 'program evaluation is a multi-disciplinary and unfortunately largely atheoretical activity'. Chen (1990) suggests that this atheoretical view tends to result in a simple input/output (or black box) type of evaluation, characterised by adherence to a step-by-step cookbook method of doing evaluation. He argues that such simple evaluations may provide evidence as to whether a program works or not, but fail to identify the underlying mechanisms that generate the treatment effects, and hence to pinpoint the deficiencies of the program for future program improvement. Therefore, theory-driven approaches offer an important contribution to good evaluation: that evaluation should move beyond merely asking why a program works, but look specifically at what aspects of a program work in which situations and why.

Critics of this approach, such as Scriven (1994), disagree with the notion that understanding the underlying theory of a project is critical and state that program theory is not needed for evaluation – as it is not needed to determine whether something works or not:

In program evaluation, we are concerned to establish the merit, worth, quality or value of programs... We do not need to know how the programs work or why they fail to work, or even what their components are (Scriven, 1994:75).

Scriven (1994) contends that understanding the mechanisms behind an intervention can be advantageous, but the main necessity is to get the evaluation right, and 'jeopardising that by diversion of effort into the direction of explanation diagnosis and remediation is all too common' (ibid: 75). Indeed, it seems that in some cases, depending on the nature of the evaluation study, an investigation into the underlying mechanisms might be something of an overkill; while unpacking the black box is generally considered useful to any evaluation, it may not be necessary for all, and its relevance depends on the nature of the evaluation inquiry (Scriven, 1994).

Depending on the nature of the evaluation questions, a theory-driven approach may or may not be appropriate. Firstly, as argued by Scriven (1994), there are situations when the cost-benefit ratio of conducting an extensive theory-driven evaluation is not justifiable. Secondly, in highly participative programs there may be instances when the nature of the evaluation questions rightfully places more

emphasis on answering specific stakeholder concerns than on investigating carefully crafted analytical suppositions.

Leviton (1994) states that theory-driven approaches are most seriously deficient in assignment of value to community-based programs. She suggests that when 'experts' frame questions for the community they encounter understandable fury. This is noteworthy as many extension programs are highly participatory and community based in orientation. In many evaluation studies in extension, the evaluators would endorse the position that the clients' and key stakeholders' values are the ones that should be used to assign value. Thus, in these cases, the use of a deductive-normative theory for developing the key evaluation questions would be rejected in favour of developing key questions through the achievement of some sort of consensus from the project stakeholders with regard to what they believe should be measured in the evaluation.

Theory-driven evaluation does not focus overtly on how unintended or unanticipated outcomes are to be investigated. Because of this there is a tendency to get drawn into the micro mechanisms that have been postulated, neglecting to look at the data afresh for new emerging patterns or mechanisms. Hamilton-Smith and Hopkins (1998) suggest for example, that in Realistic Evaluation, no overt demand is made to include unintended consequences once plausible and rival explanations are developed. Theory-driven models have also been criticised as they do not focus on how the evaluation findings are to be transferred into decisions that will lead to improved programs.

Commonalities and differences between evaluation approaches

The examination of different normative models illustrates that there are many different concepts concerning 'what evaluation is supposed to do'. However, it can be said that ultimately all of the evaluation approaches have one thing in common; they attempt to bring about improved programs – that is, programs that better meet the needs that they were designed to address. In turn, these improved programs are all intended, in one way or another, to ameliorate social, economic or environmental problems. Figure 2 shows a program theory model of the different approaches to program evaluation. While each of the approaches strive to bring about improved programs, they do this by different means, and therefore have different intermediate outcomes. Other authors have presented program theory models for program evaluation including Patton (1997), Rogers (1996) and Shadish et al. (1995).

Patton (1997) suggests that evaluation should be focused on using different organising concepts. He

suggests that in addition to program goals, evaluation can be focused upon: future decisions, critical issues or concerns, key questions and the multiple perspectives of different stakeholder groups. Figure 2 summarises the central organising concepts in each of the six normative approaches to evaluation, as well as the outcomes and intermediate outcomes for each.

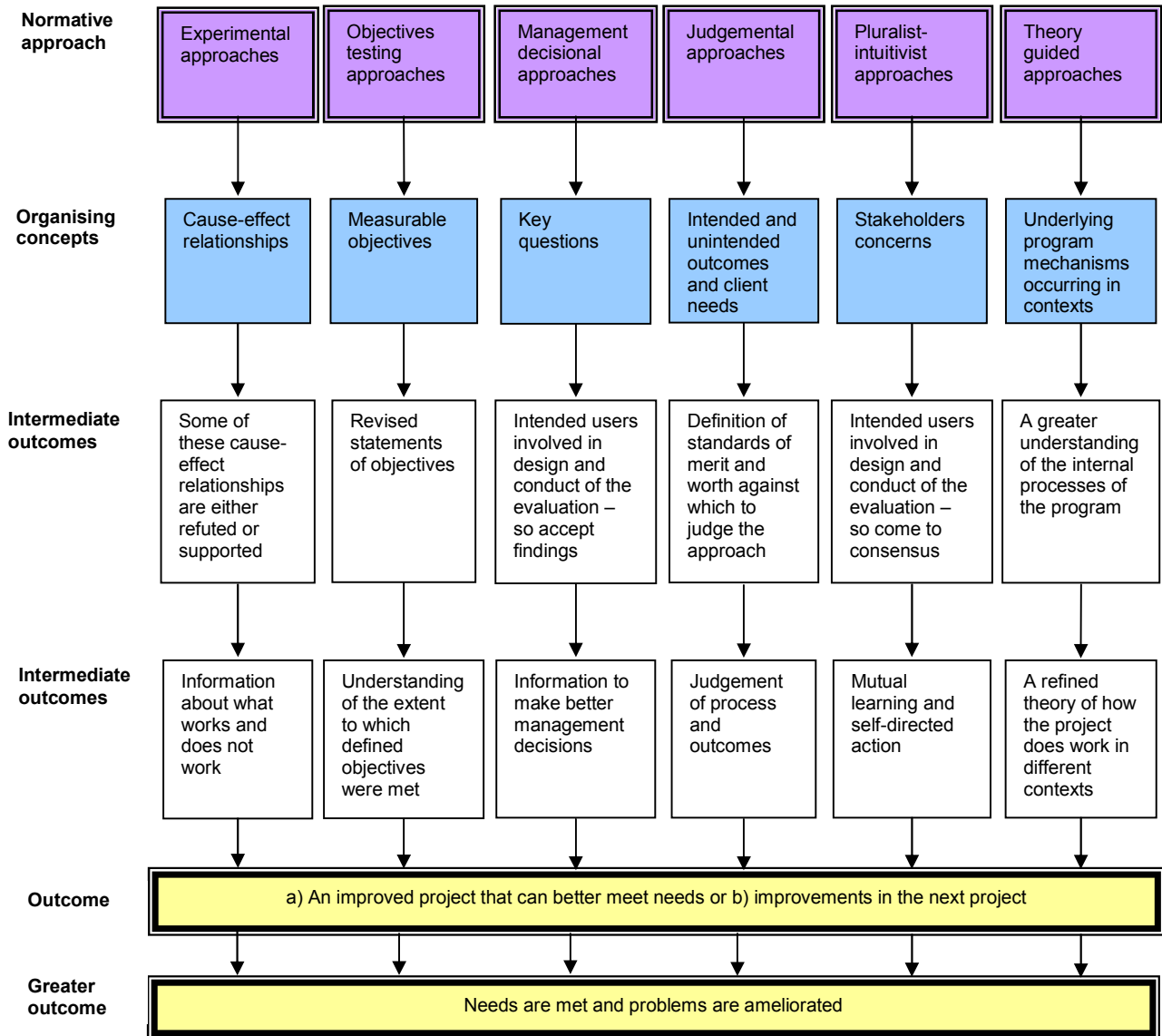


Figure 1 Outcomes and intermediate outcomes of six different normative approaches to evaluation

The central differences between the six normative approaches reviewed, and summarised in Figure 2, give rise to the following questions:

- Which organising concepts should an evaluation be built around – goals, unexpected outcomes, concerns and issues, or intended use?
- To what extent should the evaluation include judgement by the evaluator? What does this judgement consist of? If not the evaluator, who should be making the judgements?
- What efforts should be made to ensure that the evaluation findings are used?
- Who should the evaluation serve – society, the evaluation client, or the users of the service?
- To what extent should the key stakeholders be involved in the conduct and design of the evaluation?
- Who should decide on the key evaluation questions or performance indicators, ie: who decides what is measured?
- To what extent should evaluation be guided by the underpinning theory of the program intervention?

These questions are not easily answered, yet the answers for each specific evaluation study may guide the would-be evaluator toward a particular model of evaluation and away from others. I suggest that within one large extension project it may be necessary to adopt more than one of these normative approaches, entirely or partially, to meet the evaluation demands of stakeholders and to lead to improved projects. Because of this, I suggest that a pragmatic approach be adopted, whereby not only methods, but choice of normative approaches used to guide an evaluation be considered in the light of the purpose of the evaluation study. In other words, the above questions could all be answered with the statement ‘it depends on what you are trying to do’. This view corresponds with that of Kaplan (1964) who suggests that the emphasis must be on making the methodology fit the needs of the society, its institutions and its citizens, rather than the reverse.

Therefore, I advocate a framework for picking and choosing between different evaluation models. ‘Picking and choosing’ between different evaluation models can signify one of two things: choosing one evaluation model in particular to guide the evaluation or choosing bits of different models, to develop a ‘tailor-made’ evaluation. Either of the two approaches can be used.