

A new typology of monitoring and evaluation approaches

by

Babette Rabie and Fanie Cloete

ABSTRACT

The continual development of new evaluation approaches informing the evaluation of public policies and programmes, necessitates the re-grouping and re-classification of approaches with similar focus in a new typology to make sense of the different approaches and the linkages between classic and contemporary approaches. The article is based on a review of previous authors' attempts in classifying the most authoritative evaluation theories and theorists' approaches. The aim of the article is an improved classification system which includes contemporary approaches and encourages continuous thought and debate on how best to approach evaluation. The approach that is followed is therefore a theoretical re-conceptualisation approach after a critical qualitative assessment of the current literature on the topic. This paper starts by summarising the influences of the policy and social sciences on the evaluation profession. It then provides a brief overview of alternative classification systems of evaluation approaches, before suggesting a new classification system based on three categories, namely scope, focused on interventions or parts thereof; philosophy, sub-divided into evaluation studies that focus on clarifying the intervention theory and those that emphasise participation, development and empowerment; and methodology, sub-divided into experimental and non-experimental evaluation studies. Various evaluation approaches within each of the three categories are also identified and summarised.

INTRODUCTION

The evaluation or determination of the relative worth of something must be undertaken in order to compare alternatives before make choices amongst them. Evaluation literally means ‘to work out the value (of something)’ in its Latin root ‘valére’ (Mark, Greene & Shaw 2006:6). Informal evaluations inform daily decisions on how good or bad, desirable or undesirable something is. Formal evaluations involve the same kind of judgement, but are more systematic and rigorous than their informal counterparts, with appropriate controls for the validity and reliability of the findings and conclusions.

The evaluation profession emerged from various disciplines during the first half of the 20th century. Dominated by an American perspective, studies on the origins of public programme evaluation cite Ralph Tyler’s educational program evaluation (Alkin & Christie 2004:17-18) with the juvenile delinquency programme enacted by congress in 1962 as the first federally funded evaluation study (Weiss in Shadish, Cook & Leviton 1991:25). However, the systematic evaluation of public programmes already started towards the end of the Second World War (1943) in the United States (Auer 2007:541; deLeon & Vogenbeck 2007:504). The vast expansion of government social programmes throughout the 20th century is regarded as one of the critical drivers of programme evaluation (deLeon & Vogenbeck 2007:519; Shadish & Luellen in Mathison 2005:184). The evaluation field exploded in the 1960’s and 1970’s with the expansion of social policies and programmes aimed at affecting various normative and empirical goals to promote socio-economic development. During this period numerous evaluations were performed in response to federal, state and local programme managers mandates. Cost constraints and a concern about the success of social programmes in achieving outcomes fuelled the evaluation profession (Shadish & Luellen in Mathison 2005:185; Shadish, Cook & Leviton 1991:22). Chelimsky confirms that the main aim of evaluation efforts was to rationalise resource allocation and the management of programmes (2006:34). The 1970’s was marked by an increasing resistance to the expansion of social development programmes, partly as a result of the increased funding needed to sustain these programmes and the apparent ineffectiveness of many initiatives (Freeman & Solomon in Rossi, Lipsey & Freeman 2004:14). The 1980’s saw a

decline in evaluation activities under the budget cuts of the Reagan administration (Cronbach in Shadish, Cook & Leviton 1991:27). By the 1990's, fiscal and social conservatism started to thwart further expansion of government programmes, leading also to a decline in funding available for evaluation studies. (Shadish & Luellen in Mathison 2005:186).

Two specialised social science disciplines largely influenced the evaluation profession, namely public policy analysis and general social research approaches and methods. The policy analysis field is characterised by a shift from opinion-driven policy choices, to evidence-influenced and -based policy-making that explicitly takes normative decision drivers to resolve societal problems into account, while new social research theories, methodologies and technologies in the social research field brought greater research efficiency, effectiveness and sophistication to the field of evaluation research. A brief discussion of these two paradigmatic changes and their respective influences on the evaluation profession is provided below.

EVIDENCE-BASED POLICY-MAKING

The concept of 'the policy sciences' was crystallized by Harold D. Lasswell in 1943, who, together with his collaborator Lerner, re-conceptualised the intellectual tools needed to support problem-oriented, contextual, and multi-disciplinary and explicitly normative inquiry to promote equitable development (Brunner 1997:191; de Leon & Vogenbeck 2007: 509; Auer 2007). This new development was a direct response to the horrors of the Second World War, in order to avoid future repetition of these tragedies. The main foci of the policy sciences developed from the design of social reconstruction programmes for Europe and later African and other colonies in the 1950s, to an explicit general evaluation focus of those programmes during the 1960s, a further shift to specific cost-benefit assessments (Nagel & Teasley 1998) and the improvement of policy programme implementation & termination in the 1970s, better post-Vietnam & Watergate democratic ethics & values in the 1980s, public choice, business process re-engineering and new public management in the 1990s, and the current emerging focus on governance, governing networks, complex systems & technology to improve programme outcomes and impacts in the first decade of the 21st century (Auer 2007: 554; deLeon & Vogenbeck 2007: 506).

Public policy research in the second half of the 20th century thus shifted from “sterile academic parlor games” to become problem and solution oriented, focused on the “real world”. It is committed to particular values (deLeon & Vogenbeck 2007:514), “thus avoiding the value neutrality stance that social science ought to be totally objective” (Ascher 1986:365) and emphasised that the search for solutions to problems should not be lost in “scientific analysis” (Ascher 1986:370). Public policy analysis thus became more “than simply addressing big theoretical questions” but encompassed the difficult task of “clarifying goals, trends, conditions, projections and alternatives” within the social environment (Ascher 1986:371). Segone identifies an emerging shift from ‘opinion-based’ policy practice that relied either on selective evidence or on untested views of individuals or groups, through ‘evidence-influenced’ to ‘evidence-based’ policy practices that place the best available evidence from research at the heart of policy development and implementation to improve policies (Segone 2009:17; Davies, Newcomer & Soydan 2006:175; Davies, 2008:3).

Segone attributes the emergent shift from opinion-based to evidence-influenced approaches to movements towards more transparent governance and better technical capacity to produce quality, trustworthy evidence (Segone 2009:18). This shift is also driven by international initiatives like the Millennium Development Goals Programme, World Bank initiatives, Transparency International and donor funding that emphasise the need to evaluate the success of public policies and programmes in order to protect donors’ investment in these programmes (Kusek & Rist 2004: 3-11; Valadez & Bamberger 1994:5-7). Internal fiscal constraints, pressures for public accountability and the failure of past programmes to produce results also emphasise evaluation to ensure resources are allocated to the most pressing problems and the most effective and efficient programmes to address these problems (Rossi et al 2004:15; Boyle & Lemaire 1999:3&181). Ideally, however, policy practices should be evidence-based, but this implies depoliticising political discretions to a bigger extent to curb decision drivers like hidden agendas and other subjective priorities of decision-makers that are not necessarily evidence-based. This outcome is not attainable in current political decision-making systems.

SOCIAL RESEARCH

Up to the end of the Second World War, social research comprised mainly of qualitative research designs and methods because of a lack of adequate quantitative empirical methodologies and technologies to analyse and assess big data sets with the required scientific rigour. While education and health programme evaluation studies were already undertaken since the mid eighteenth century, programme evaluation only attained the minimum necessary level of measurement, sampling and statistical sophistication during the fifties and sixties in order to realise the hope that social science research could mimic the success of physical science research in solving technological problems in the social arena (Mouton 2007:492; Shadish & Luellen in Mathison 2005:184). One of the most influential stimuli in this regard was Campbell and Stanley's 1966 paper on '*Experimental and Quasi-Experimental Designs for Research*' that reformed social science research by emphasising experimental design and randomisation, internal and external validity and alternative 'quasi-experimental' designs (Alkin & Christie 2004:19-20). Suchman's 'Evaluative Research' (1967) applied these social research methods to evaluation and signified the birth of evaluation research as an applied social research activity (Alkin & Christie 2004:22-25).

The second major innovation that enabled the increasing significance of quantitative approaches to programme evaluation is the development and maturity of the information society during the last decades of the 20th century, mainly as a result of the increasing sophistication of computer technologies. The development of the personal computer and increasingly sophisticated computer hardware and software operating systems as well as mathematical and statistical programmes enabled the mass collection, storage, processing, analysis and assessment of huge volumes of general social research data that was not possible until that time. This development facilitated the accuracy, validity, efficiency and effectiveness of general social research processes and outcomes, and also had an automatic beneficial overspill effect on more specialised policy, programme and project evaluation exercises where such research approaches and methodologies could be applied (Ayres 2008; Dubner & Levitt 2006).

In supporting the view of programme evaluation as an applied and specialised social research activity, Bickman defines evaluation research as an assessment of the strengths and weaknesses of a policy intervention, identifying ways to improve them, and determining whether desired outcomes are achieved (in Mathison 2005:141). Freeman & Rossi define evaluation research as “the systematic application of social research procedures for assessing the conceptualisation, design, implementation and utility of social intervention programmes” (in Mouton 2007:491). Scriven, however, is of the opinion that evaluation research is much more than just applied social research and that while evaluators need a repertoire of empirical research skills, they also require additional evaluative skills that enable them to search for side effects that may influence the evaluation conclusion, determine relevant technical, legal and scientific values and synthesis skills to integrate evaluative and factual information (Scriven 2003:7). This positivist conclusion by Scriven is controversial, because the question arises whether an academic social researcher should not also ideally have these skills?. Distinctions between evaluation and such a positivist view of social research can also include value neutrality versus value laden studies, a conclusion about the merit or net benefits through the verification of values and standards and a shift from mere knowledge production in research to knowledge application in evaluation (Scriven 2003:7; Rossi et al 2004:17; Weiss in Shadish, Cook & Leviton 1991:182). However, if one moves beyond this narrow behaviourist or positivist view of social research and accepts that intangible, normative variables also influence social attitudes, processes, structures and their consequences for society, then the putative distinction between social research and evaluation lapses. Evaluation has in fact moved beyond Campbell’s original behaviourist methodological focus in the 1960s to embrace more intangible concepts such as utilisation, values, context, change, learning, strategy, politics, and organisational dynamics (McClintock 2003:14).

From this wider perspective, any evaluative social research conclusion is thus a blend of fact and value claims, and it also entails discovering the right criteria and standards for comparison with other social phenomena to make sufficiently rigorous generalisations to improve theory and practice (House 2004:219). From this perspective evaluation is nothing more than an applied social research activity.

EVALUATION APPROACHES: TOWARDS AN IMPROVED CLASSIFICATION SYSTEM

During its relatively short history, the evaluation profession has already been characterised by a variety of philosophies, approaches, models, traditions and practices. Classification schemes are usually criticised on the basis of what they include and exclude. However, the charting of evaluation approaches has a pragmatic purpose as it provides evaluation practitioners with the detail to make a choice amongst various evaluation approaches based on their inherent parameters, purposes and processes (Mathison 2005:257). It allows for the most appropriate fit between the evaluation exercise's purpose, its underlying values and the most appropriate methodologies to achieve the most rigorous results.

The first evaluation studies tested bold new reform approaches, while ignoring the effects of small changes to existing programmes or local practices for local goals. Over time evaluation approaches changed and diversified to reflect accumulating practical experience (Shadish, Cook & Leviton 1991:32). While early theories focused on methods for doing evaluations in natural field settings, later theories focused on the politics of applying methods in field settings, and how research fits into social policy (Shadish & Luellen in Mathison 2005:186).

Various attempts have been made to classify these alternative competing theories and models, signalling a natural growth in the evaluation discipline to assist better evaluation theory and practice (Mathison 2005:258). While every attempt adds value to the evaluation field in broadening understanding of the similarities and differences of various approaches to evaluation, none of the current classification attempts seem to cover the wide range of approaches that are currently in existence adequately and comprehensively enough. This might create confusion among scholars and practitioners to accurately understand the phenomena that they are dealing with.

Shadish, Cook and Leviton (1991) for example classify theories and theorists into three 'stages': stage one theories which introduce science and experiments as means to address social problems, stage two theories and theorists that emphasise use and pragmatism, and stage three theories which try to integrate the scientific

and pragmatic approaches. The problem with this classification is that it is a formalistic classification of the historical development of evaluation foci that only distinguishes between two variables, namely more rigorous methodologies for evaluation on the one hand versus less rigorous methodologies with a stronger focus on societal relevance and use on the other hand. This approach is not very useful because it does not distinguish between potentially different competing normative goals for the evaluation exercise and the scope or degree of comprehensiveness of the evaluation which has an important implication for decisions about the timeframes, resources and methodologies to be employed in the evaluation project.

Chen presents four types of evaluation strategies linked to the purpose of the evaluation. He distinguishes between evaluation strategy (the general direction taken by the evaluator to meet a particular purpose) and evaluation approach (the systematic set of procedures and principles guiding evaluators, including conceptualising problems, research method application and interpretation of data) (2005:144). Four types of strategies are identified, namely assessment strategies that provide information on the performance of the intervention, development strategies that assist in planning the intervention, enlightenment strategies that examine underlying assumptions and mechanisms to mediate observed effects, and partnership strategies that involve stakeholders in planning and implementing interventions (Chen 2005:144-148). While this approach provides a useful classification in terms of the objective that an evaluation study may pursue, it omits the focus and methodology of the evaluation study from the taxonomy, although Chen does address this peripherally in his steps to apply the taxonomy. Furthermore, while Chen's classification taxonomy describes the broad aim of each strategy, it fails to provide examples of approaches of other authors within the four aims which may clarify the issue better.

Alkin and Christie developed an evaluation tree with three main branches, namely use, methods and valuing (2004:12). Various evaluation theorists were sorted onto the three branches based on their (most) important contributions to the evaluation field. However, many theorists' contributions, especially those who have grown over time in the field and changed their viewpoint, span across the branches, resulting in a forced classification onto one branch that does not provide an accurate reflection

of their contributions to the evaluation field. A focus on approaches, rather than authors, allow for overlapping and changing views.

Rossi et al (2004:40) and Mouton's (2008) classification systems link evaluation to the programme life cycle (design, implementation and outcomes). Similarly, Owen (2006:41-54) distinguishes between proactive evaluation aimed at synthesising previous evaluation findings, clarificative evaluation to clarify the underlying logic and intended outcomes of the intervention, interactive evaluation to improve the evaluation design, monitoring evaluation to track progress and refine the programme and finally impact evaluation for learning and accountability purposes. While this typology is useful in identifying the design, tracking implementation and evaluating final results, it ignores other variables and choices that need to be taken into consideration in deciding what approach to be followed during the evaluation exercise.

Stufflebeam identified a comprehensive shopping list of 26 approaches to evaluation classified into five categories: Pseudo-evaluations, Questions- and Methods-Oriented Evaluation Approaches (Quasi-Evaluation Studies), Improvement- and Accountability-Oriented Evaluation Approaches, Social Agenda and Advocacy Approaches, and Eclectic Evaluation Approaches (Stufflebeam & Shinkfield 2007). This is the most recent and comprehensive attempt at systematising evaluation approaches, and clearly Stufflebeam and Shinkfield has undertaken an exhaustive assessment of different evaluation approaches. Their proposed classification system is more useful than any of the previous attempts, but can still be refined further, because the 'shopping list' nature of the classification system makes it very unwieldy and clumsy, and still contains too many overlapping approaches.

Our alternative classification system proposed below attempts to supplement the identified weaknesses of the above classification schemes. It uses three main classification categories, namely the **scope** of the evaluation study, the approach or underpinning **philosophy** of the evaluation study, and lastly the evaluation study **design** and methodology which provides the parameters for collecting and assessing data to inform the evaluation. We have tried to build on the strong elements of the various classifications summarised above and added more

dimensions that we regard as relevant and useful. The focus of our proposed model is to provide a more accurate combination of **parameters**, implicit or explicit **normative or value frameworks** underlying the evaluation exercise and alternative **designs and methodologies** for evaluation. The scope of the study defines the parameters of the evaluand. The particular objectives of the study informs the choice of philosophy or evaluation approach, and while “the various approaches to evaluation are all defensible, (they are) not necessarily equally defensible in any given evaluation situation” (Rossi et al 2004:26). Finally, the specific evaluation designs and methodologies employed, evaluation question(s) and data sources provide for the selection of appropriate data collection, analysis and assessment methods.

The main evaluation approaches that emerged during the last 50 years of evaluation research will be discussed below within these three categories of the proposed new classification system.

Evaluation approaches based on scope

The functional, geographic or behavioural parameters of the evaluation determine and delimit the focus of the evaluation. The evaluation may be very broad, encompassing several of the dimensions or attributes of performance listed below, as is done during a comprehensive organisational performance review. A comprehensive evaluation therefore focuses on more than one and even in extreme cases on all of the aspects of the evaluation (integrated evaluation). Alternatively, the evaluation may be focused on a particular intervention, be that a policy, a programme, a project or a product; or limited to a particular development sector (eg the economy, political, financial, technological, cultural, environmental, educational, transport, health or other sectors of a community or society), geographical area or community; confined to a particular phase or stage of an intervention (such as its inputs, resource conversion or management processes, outputs, outcomes or impacts); or focused on the performance of individual staff members within the organisation or intervention. Only organisational evaluations are, however, dealt with below.

The main evaluation approaches based on scope, are the following:

- **Community-based evaluation** focuses on a particular community, which may be geographically based, or spatially spread, but with similar characteristics such as ethnicity, interest or ideology (Conner in Mathison 2005:69-70).
- **Sectoral evaluations** evaluate different sectoral policies, programmes and/or projects like transport, education, health and welfare.
- **Geographical evaluations** evaluate the consequences of specific location-based policies, programmes and/or projects like integrated community, local government, regional, provincial or national developmental initiatives.
- **Policy evaluation** can focus on either policy process assessment (how and why policies are devised and implemented) or policy content assessment (what interventions are considered or made), or both (Owen 2006:26).
- **Programme and project evaluation** systematically investigates the effectiveness of social intervention programmes/projects in ways that are adapted to their political and organisational environments to inform social actions that may improve social conditions. (Rossi et al 2004:431). Program evaluation also assesses the program results and the extent to which the program caused those results" (Wholey, Hatry & Newcomer 2004:xxxiii).
- **Product evaluation** entails the evaluation of (not the process, but) only the product against quality assurance standards. In the social context, product evaluation measures, interprets, and judges the achievements to ascertain the extent to which the evaluand met the needs of the rightful beneficiaries (Stufflebeam & Shinkfield 2007:344-345).
- **Input evaluation** assesses only the required financial, human, physical, time, information and commitment resources. It enables decision makers to examine the feasibility of alternative strategies for addressing identified needs of targeted beneficiaries to prevent failure or waste of resources (Stufflebeam 2004:338-339).
- **Process or ongoing evaluation** investigates only the implementation of the programme, including whether the administrative and service objectives of the programme are being met; whether services are delivered in accordance to the goals of the program; whether services are delivered to appropriate

recipients and whether eligible persons are omitted from the delivered service; whether clients are satisfied; whether the administrative, organisational and personnel functions are managed well; whether service delivery is well-organised and in line with programme design and other specifications and whether the project runs within the projected budgetary and time frames (Rossi et al 2004:56-57, 78, 171).

- **Output evaluation** assesses the tangible product or service produced by the intervention in terms of the quantity, quality and diversity of services delivered. It is the easiest and most straight-forward focus for evaluation.
- **Outcome evaluation** focuses on the positive, neutral or negative intermediate sectoral results or consequences of a project/programme (ie progress made towards achieving the strategic goals) (Rossi et al 2004:224-225; Chen 2005:35; Weiss 1998:8). Outcome evaluations may focus on “the individual level (changes in knowledge, skills, attitudes), organisational level (changes in policies, practices, capacity), community level (changes in employment rates, school achievement, recycling), and the policy or government level (changes in laws, regulations, sources of funding)” (Mathison 2005:287).
- **Impact evaluation or impact assessment** focuses on final long term multi-sectoral consequences of the project/programme (ie progress towards achieving the transformative vision). It determines “the extent to which a program produces the intended improvements in the social conditions it addresses” (Weiss, 1998:8). It tests whether the desired effects of the social conditions that the programme intended to change, were attained and whether those changes included unintended side effects (Rossi et al 2004:58, 427; Owen 2006:255).
- **Systemic evaluation** analyses the entire system, including the effect of external factors on the system, with the aim of improving its functioning (Rogers & Williams 2006:88).
- **Integrated evaluation** such as Stufflebeam's CIPP approach, logic models and logical frameworks combines the various focus areas into an overall assessment.

- **Meta-evaluation** evaluates the evaluation focus, content and process as well as the evaluators themselves (Scriven in Mathison 2005:249-251). Interpretations by evaluators and others should be scrutinized by colleagues and selected stakeholders... to identify shortcomings in design and poor interpretations" (Stufflebeam interpreted by Stake 2004:215).

Evaluation approaches based on an explicit philosophy or formal substantive theory

The various philosophical/theoretical approaches to evaluation range from “a largely positivistic perspective on the one hand where quantitative approaches are used to generate clinical information about measurable and calculable behaviour patterns analysed on the basis of so-called scientific criteria (eg the analysis of huge quantitative datasets), to the more normative, interpretative and constructivist approaches on the other hand which prioritise the identification and generation of local knowledge, learning and use within the context of different situations and cultures (eg the assessments of similarities and differences between specific case studies). This broad distinction of the two polar opposites of approaches in this category classifies adherents into two camps, the quantitative or ‘scientific’ versus the qualitative or interpretative, the former being seen as closer to the more rigorous natural sciences and the latter closer to the more ‘fuzzy’ social sciences” (Naidoo, 2007:31).

Some of the previous evaluation approach classification attempts distinguish between value-driven and use-driven evaluation approaches. The problem with this distinction is that all evaluations inherently entail a value judgement (good or bad, in Scriven’s simple distinction) and that all evaluations are goal-directed with a particular end-use or purpose in mind. A clearer distinction in terms of the underpinning philosophy of an evaluation is theory-driven versus participation-driven approaches, where theory-driven evaluation philosophies lean towards a more scientific approach to evaluation research with the general aim to expand knowledge, while participation-driven evaluation philosophies lean towards a more applied social improvement approach to evaluation research with the general aim of

development, empowerment and creating shared understanding of the programme between the evaluators, beneficiaries and decision-makers.

Theory-based evaluation entails the identification of the critical success factors of the evaluation, linked to an in-depth understanding of the workings of a programme or activity (the “programme theory” or “programme logic.” Summarised in the previous category). Theory-driven evaluation is therefore ‘the systematic use of substantive knowledge about the phenomena under investigation and scientific methods to improve, to produce knowledge and feedback about, and to determine the merit, worth and significance of evaluands” (Donaldson & Lipsey 2006:67) (eg assessing sectoral or integrated governmental interventions to reduce poverty, unemployment, crime and insecurity, and to improve health, education, quality of life and community development). The approaches in this category are all based on an implicit ‘theory of change’ (eg how to reduce crime, poverty and disease and achieve growth and development), which links the evaluation with intended improvements in practice (Rogers & Williams 2006:77). It does not assume simple linear cause-and effect relationships, but allows for the mapping and design of complex programmes. Where evaluation data indicates that critical success factors of a programme have not been achieved, it is concluded that the programme will be less likely to succeed (Kusek & Rist 2004:10).

These evaluations can be approached in a deductive or an inductive way. The following are specialised **deductive** approaches:

- **Clarification evaluation**, or the assessment of programme theory (about the above examples and others), assists to clarify or develop the programme plan (Chen 2005:127); to analyse the programme assumptions and theory (Rossi et al 2004:93); to determine its reasonability, feasibility, ethics and appropriateness (Rossi et al 2004:55), and improve coherence (Owen, 2006:191). Clarificatory evaluation tests the deductive or inductive causal logic of the intervention, the feasibility of the design, encourages consistency between design and implementation (Owen, 2006:192). A useful approach is to draw the causal “logic model” for the intervention to provide a picture of how it is believed the

intervention will work to bring about desired results through a specific sequence of activities (Kellogg Foundation 2004:10).

- **Illuminative evaluation** is basically the same as clarification evaluation. It assesses the significant features, recurring issues and themes and critical processes of a programme to provide a comprehensive understanding of the complex reality surrounding a program: in short, to 'illuminate' (Hamilton in Mathison 2005:191-194). In contrast to clarification evaluation which is a deductive approach from within the perspective of a specific theoretical paradigm, illuminative evaluation, however, generally follows an inductive approach.
- **Realist evaluation** tries to establish why, where, and for whom programmes work or fail by identifying the mechanisms that produce observable programme effects. It can also test the mechanisms as well as other contextual factors that may have caused the observed effect (Henry in Mathison 2005:359). It thus tests whether there is an unequivocal causal relationship between a programme and its outcomes to establish beyond doubt that it was the actual programme which caused the measurable change, and not some other, unidentified, variable which may not exist in another social setting (Mouton 2008).
- **Cluster evaluation and multisite evaluations** look across a group of projects to identify common threads and themes across such projects (Russon in Mathison 2005:66-67). Cluster evaluation tries to establish impact through aggregating outcomes from multiple sites or projects, whereas multisite evaluation seeks to determine outcomes through aggregating indicators from multiple sites. Both approaches try to clarify and verify the validity of the theory of change concerned.

Goal-free evaluation is an example of an **inductive** theory-driven approach:

- **Goal-free evaluation** studies all aspects of the programme and notes all positive and negative aspects without focusing only on information that supports the goals (Posavac and Carey 1997:23-27). The evaluator remains purposely ignorant of a program's goals, searching for all effects of a program regardless of its developer's objectives. If the program is doing what it is supposed to do, the evaluation should confirm this, but the evaluator will also be more likely to uncover unanticipated effects that the goal-based evaluations would miss because of the preoccupation with stated goals" (Stufflebeam & Shinkfield

2007:374). Conceptualised in this way, goal-free evaluation is seen as the opposite of a **deductive** theory-driven approach to evaluation.

“Participatory evaluation is an overarching term for any evaluation approach that involves program staff or participants actively in decision making and other activities related to the planning and implementation of evaluation studies” (King in Mathison 2005:291-294). In participatory evaluation the evaluation team consists of the evaluator (either as team leader, or as supportive consultant) and representatives from stakeholder groups, who together plan, conduct and analyse the evaluation. The degree of participation can range from shared evaluator-participant responsibility for evaluation questions and activities, to participants' complete control of the evaluation process. With shared responsibility, the evaluator is responsible for the quality of the process and the outcomes, but designing and conducting the evaluation is done in collaboration with stakeholders. In evaluations where participants control the evaluation, the evaluator becomes a coach or facilitator who offers technical skills where needed. In a sense, all evaluations have some participation from stakeholders as evaluators need to interact with stakeholders to obtain information. However, a study has a participatory philosophy when the relationship between the evaluator and the participants provides participants with a substantial role in making decisions about the evaluation process.

- **Responsive evaluation** is not particularly responsive to program theory or stated goals but more to stakeholder concerns (Stake & Abma in Mathison 2005:376-379). In contrast to pre-ordinate goal-focused evaluation where the evaluator predetermines the evaluation plan, based on the programme goals, responsive evaluation orients the evaluation to the programme activities as oppose to the goals, thereby responding to various information needs and values with appropriate methods that emerge during the course of the programme implementation (Stake in Shadish, Cook & Leviton 1991:270). Responsive evaluation searches for pertinent issues and questions throughout the study and attempts to respond in a timely manner by collecting and reporting useful information, even if the need for such information had not been anticipated at the start of the study (Stufflebeam & Shinkfield 2007:415).

- **Naturalistic, constructivist, interpretivist or fourth-generation evaluation** attempts to blend the evaluation process into the lives of the people involved by focusing on both the tangible, countable reality and the intangible socially-constructed reality (what people believe to be real (Lincoln & Guba, 2004:228). “The merit or worth of the evaluand is judged in ways appropriate to the setting, expectations, values, assumptions, and dispositions of the participants, with minimal medications due to the inquiry processes used and assumptions held by the evaluator” (Williams in Mathison 2005:271). Values are assigned a central role in the evaluation, as they provide the basis for determining merit. The values of stakeholders, values inherent to the context or setting of the situation and conflict in values are critical in formulating judgements and conclusions about the evaluand (see Lincoln in Mathison 2005:161-164).
- **Utilisation-focused evaluation** is based on the premise that evaluations should be judged by their utility and actual use. Therefore, evaluators should facilitate the evaluation process and design the evaluation with careful consideration of how everything that is done, from beginning to end, will affect its use (Patton in Mouton 2007:504). A group of representative stakeholders clarify the outcomes, indicators, performance targets, data collection plan and intended uses of the findings will be used. The group’s values (not the evaluator’s) thus determine the nature of recommendations arising from the evaluation (Stufflebeam & Shinkfield 2007:434, 440). Patton argues that as evaluation cannot be value-free, “utilisation-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have the responsibility to apply evaluation finding and implement recommendations” (Patton 2004:277).
- **Appreciative inquiry** focuses on the strengths of a particular organisation or intervention with the assumption that focusing attention on the strengths will strengthen them further. Appreciative inquiry is based on the social constructivist concept that “what you look for is what you will find, and where you think you are going is where you will end off” (McClintock 2003:15).
- **Evaluative inquiry** responds to a range of decision-makers’ information needs, of which determining the worth of the programme may be one (Owen 2006:17). “Evaluative inquiry consists of collecting data, including relevant variables and

standards, resolving inconsistencies in the values, clarifying misunderstandings and misrepresentations, rectifying false facts and factual assumptions, distinguishing between wants and needs, identifying all relevant dimensions of merit, finding appropriate measures of these dimensions, weighing the dimensions, validating the standards, and arriving at an evaluative conclusion” (House in Owen 2006:17). It emphasises the importance of individual, team and organisational learning as a result of participating in the evaluation process (Mathison 2005:201).

- **Critical theory evaluation** aims to determine the merit, worth or value of something by unveiling false culturally based perspectives through a process of systematic inquiry (Greene 2006:129). “The evaluation is influenced by an explicit value position that we operate beneath layers of false consciousness contribute to our own and others’ exploitation and oppressions....As a response, critical theory evaluation seeks to engage evaluation participants in a dialectic process of questioning the history of their ideas and thinking about how privileged narrative of the past and present will influence future value judgements: (MacNeil in Mathison 2005:92-94).
- **Empowerment evaluation** uses the evaluation process to foster self-determination with the help of the evaluator coach or critical friend. The evaluator helps the group to determine their mission, take stock through evaluation tools of the current reality and to set goals and strategies based on the self-assessment (Fetterman 2004:305). The evaluator needs to capacitate stakeholders to enable them to conduct independent evaluations, thereby altering the balance of power in program context by enhancing the influence of stakeholders (Rossi et al 2004:51).
- **Democratic evaluation** considers all relevant interests, values, and perspectives to arrive at conclusions that are impartial to values (House & Howe (1999) and House 2004:220). Democratic evaluation allows the multiple reality of a program to be portrayed, providing decision-makers with a variety of perspectives and judgements to consider (MacDonald 1979 in Alkin & Christie 2004:40). House (1991 1993) argues that “evaluation is never value neutral; it should tilt in the direction of social justice by specifically addressing the needs and interests of the powerless” thereby promoting social justice to the poor and marginalised through

the evaluation process (Alkin & Christie 2004:41). Evaluation thus becomes a democratising force with evaluators advocating on behalf of disempowered groups (Mouton 2007:502).

Evaluation design and methodology

Advances in social research methods since the 1950's presents the evaluation field with various options in designing studies to collect and analyse data that informs the evaluation process. Evaluation research studies may adopt either a **quantitative, a qualitative or mixed-methods design approach**, as the evaluator tries to find a workable balance between the emphasis placed on procedures that ensure the validity of findings and those that make findings timely, meaningful, and useful to consumers. Where that balancing point will be will depend on the purposes of the evaluation, the nature of the program, and the political or decision-making context (Rossi et al 2004:25). Rossi refers to this as the "good-enough" rule, which entails choosing the best possible design, taking into account practicality and feasibility (paraphrased by Shadish, Cook & Leviton 1991:377).

While a particular evaluation approach such as the classic experimental study may be ideal, it may not be feasible. Lee Cronbach concluded in 1982 that "evaluation studies should be judged primarily by its contribution to public thinking and to the quality of service provided subsequent to the evaluation... An evaluation should inform and improve the operations of the social system with timeous feedback (not necessarily perfect information)" (Rossi et al 2004:23-24). Given the advantages and disadvantages of different approaches, the OECD argues for "the use of a plurality of approaches that are able to gain from the complementarities in the information they can provide" (OECD 2007:24). Different research methodologies may be applied in different evaluation designs. Only the main evaluation designs that are applicable to evaluation studies are summarised below. This might exclude more technical research methodologies.

Quantitative evaluation approaches normally take the form of experimental designs. An **experimental design** advocates "a social experimental approach to reform where social programmes are retained, imitated, modified or discarded on the

basis of apparent effectiveness on the multiple imperfect criteria available" (Rossi et al 2004:23-24). When a clear statement of the program objective to be evaluated has been formulated, the evaluation may be viewed as a study of change. The program to be evaluated constitutes the causal or independent variable, and the desired change is similar to the effect or dependent variable...the project may be formulated in terms of a series of hypotheses that state that activities A,B and C will produce results X, Y and Z (Stufflebeam & Shinkfield 2007:277,281).

- **Classic experimental design** entails the random assignment of subjects to treatment and non-treatment conditions, and the pre- and post measurement of both groups. The impact of programmes is determined by comparing the outcomes of the groups to determine whether the intervention has produced the desired outcome (Mouton 2007:495; OECD 2007:22).
- **Quasi-experimental evaluation** attempts to overcome the problems with randomly assigning participants to interventions in real life – as opposed to laboratory conditions (Mouton, 2007:495). The term 'quasi-experimental' refers to approximations of randomised experiments and while their control of internal validity is not as reliable as true experimental design, they nevertheless provide valuable answers to cause-and-effect questions (Campbell in Shadish, Cook & Leviton 1991:120; Mark & Henry 2006:323). The validity of the quasi-experiment may be undermined by historical or seasonal events, maturation of the subjects, the effect of the test or instruments used on the subject's behaviour, attrition of subjects from the programme and statistical regression that would have occurred naturally without any intervention (Reichardt & Mark 2004:128-129). Forms of quasi-experimental designs include pretest-posttest non-equivalent comparison group design, pretest-posttest no comparison group design, interrupted time-series designs, comparison group designs, and regression-discontinuity design where the conditions for being part of the experimental group is known and therefore 'controllable' (see Reichardt & Mark 2004).

Qualitative evaluation approaches are non-experimental approaches. They focus on the constructed nature of social programmes, the contextuality of social

interventions and importance of focusing on processes of implementation, in addition to assessing programme outcomes and effects (Mouton 2008). “Understanding the quality of the program requires understanding program activities in considerable detail. The measurement of outcomes and impact ... is often simplistic and of low validity” (Cronbach interpreted by Stake 2004:215). **Qualitative evaluation** answers ‘why’ and ‘how’ questions (Wholey 2004:269). It is ideal when non-causal questions form the basis for the evaluation, when contextual knowledge, perspective and values of the evaluand is required before finalising the evaluation design, when the focus is on implementation rather than outcomes, when the purpose of the evaluation is formative, when it is important to study the intervention in its natural setting by means of unobtrusive measures (Pierre 2004:151; Mouton 2007: 497).

- **Case study evaluation** approaches see the evaluator analysing the goals, plans, resources, needs and problems of the case in its natural setting (as opposed to imposed experimental conditions) to prepare an in-depth report on the case, with descriptive and judgemental information, perceptions of various stakeholders and experts, and summary conclusions (Stufflebeam & Shinkfield 2007:309-310). In the case study approach, the evaluator seek patterns of data to develop issues, triangulates key observations and bases for interpretation, selects alternative interpretations to pursue and develops assertions or generalisations about the case (Stufflebeam & Shinkfield 2007:314-315). Success case method compares the experiences of successful and unsuccessful participants to identify key factors that allowed successful participants to benefit from a particular intervention (Brinkerhoff in Mathison 2005:401-402; Rogers & Williams 2006:88).
- **Participatory action research** combines the investigative research process with education of less powerful stakeholders and subsequent action on the research results. The cycle starts with observation and reflection, which leads to a plan of change to guide action. The approach is best suited to action-orientated evaluation questions. (Rogers & Williams, 2006:83,84)
- **Grounded theory** provides an open-ended evaluation design where the evaluator’s inductive sensitivity, ontology and epistemology is the preferred methodological paradigm (Auriacombe 2009). Grounded theory is particularly

helpful in goal-free evaluations where it assists in developing substantive theoretical propositions and extrapolations from the classification or coding of empirical data that might lead to theory building or change, rather than the testing of a theory as happens in a deductive theory-driven approach to evaluation.

CONCLUSION

Programme evaluation is a relatively recent applied research activity. It developed from different scholarly disciplines of which the policy sciences and social research methodology traditions had the most direct impacts on the development of this field of endeavour.

Due to the complexity of evaluation studies in practice, evaluation studies do not take “one” single approach to evaluation. Past attempts to classify different approaches to evaluation have not succeeded in accurately identifying the nature and most appropriate clustering of competing approaches. Such classification is necessary to enable evaluators to understand the different approaches to evaluation and also how they relate to, overlap or differ from one another.

The three revised categories of evaluation approaches proposed above is an attempt to improve the current state of evaluation approach classification, are not water-tight distinctions. In the summaries of these approaches it is clear how some are mutually exclusive, others overlap and many are related or complementary. Dahler-Larsen views the diversity in approaches as an asset, as it sparks constant debate and new practices to new and old problems (2006:157). It also reinforces the holistic complexity of the social phenomena that we try to understand, and the fact that our current measuring instruments are still primitive and only able to provide us with approximations of the real nature of these phenomena.

In order to get the most accurate perspective of whatever we are trying to evaluate it is necessary to consider and apply different approaches. Thus, an outcome evaluation study may take a participatory approach to clarify the multiple aims and

intended uses of the evaluation results, followed by a more theory-driven approach in the summative evaluation to determine whether the predetermined goals were reached as well as identifying potential unintended consequences. The nature of the evaluand will determine the appropriate quantitative or qualitative data gathering techniques, which will inform the design of the study in addition to the stated goals of the evaluation. As the different approaches emphasise different aspects of the evaluand, it can be argued that a combination of approaches will provide 'richer' evaluation data through a multifaceted evaluation focus. However, each additional approach implies more resources (including time) to bring it to fruition. It is the task of the evaluator to select the most appropriate balance of approaches to ensure the most accurate evaluation results within the limited resources available.

REFERENCES

- Alkin, C.M. & Christie, C.A. 2004. An Evaluation Theory Tree. In Alkin, C.M. (Ed).2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: SAGE Publications.
- Ascher, W. 1986. The Evolution of the Policy Sciences: Understanding the Rise and Avoiding the Fall. *Journal of Policy Analysis and Management*, Vol. 5, No. 2. (Winter, 1986), pp. 365-373.
- Auer, Matthew R. 2007. The Policy Sciences in Critical Perspective. Chpt 14 of Rabin, Jack, Hildreth, W B & Miller, G J (eds). *Handbook of Public Administration*. 3rd edition. New York: Taylor & Francis. 541-562.
- Ayers, Ian. 2008. *Supercrunchers: How anything can be predicted*. London. John Murray Publishers.
- Boyle, R. & Lemaire, D. (editors) 1999. *Building Effective Evaluation Capacity. Lessons from Practice*. New Brunswick. USA.Transaction Publishers.
- Brunner, R.D. 1997. Introduction to the policy sciences. *Policy Sciences* Vol 30 pp 191-215.
- Chelimsky, E. 2006. The Purposes of Evaluation in a Democratic Society. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Chen, H. 2005. *Practical Program Evaluation. Assessing and Improving Planning, Implementation and Effectiveness*. California: Sage Publications.

- Dahler-Larsen, P. 2006. Evaluation after Disenchantment. Five Issues Shaping the Role of Evaluation in Society. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Davies, P. 2008. *Making Policy Evidence-Based: The UK Experience*. World Bank Middle East and North Africa Region Regional Impact Evaluation Workshop Cairo, Egypt, 13-17 January 2008. Available online at http://siteresources.worldbank.org/INTISPMA/Resources/383704-1184250322738/3986044-1209668224716/English_EvidenceBasedPolicy_Davies_Cairo.pdf, accessed 30 April 2009.
- Davies, P., Newcomer, K. & Soydan, H. 2006. Government as Structural Context for Evaluation. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- de Leon, Peter & Vogenbeck, Danielle M. 2007. *Back to Square One: The history and Promise of the Policy Sciences*. Chpt 13 of Rabin, Jack, Hildreth, W B & Miller, G J (eds): *Handbook of Public Administration*. 3rd edition. New York. Taylor & Francis. 503-539.
- Donaldson, S.I. & Lipsey, M.W. 2006. Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Dubner, Stephen J & Levitt, Steven D. 2006. *Freakonomics*. London: Penguin Books.
- Fetterman, D.M. 2004. Branching Out or Standing on a Limb: Looking to Our Roots for Insight. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Greene, J. C. 2006. Evaluation, Democracy, and Social Change. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- House, E.R. 2004. Intellectual History of Evaluation. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Kellogg Foundation. 2004. *Using Logic Models to Bring Together Planning, Evaluation, and Action*. Logic Model Development Guide. Available online at.... Accessed on 25 November 2008

- Kusek, J.Z. & Rist, R.C. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington D.C. The World Bank.
- Lincoln, Y.S. & Guba, E.G. 2004. The Roots of Fourth Generation Evaluation: Theoretical and Methodological Origins. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Mark, M.M & Henry, G.T. 2006. Methods for Policy-Making and Knowledge Development Evaluations. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Mark, M.M., Greene J.C. and Shaw, I.F. 2006. Introduction The evaluation of policies, programs and practices. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Mathison, S. (editor) 2005. *Encyclopedia of Evaluation*. California: Sage Publications.
- McClintock, C. 2004. *Using Narrative Methods to Link Program Evaluation and Organizational Development*. The evaluation exchange. Vol IX, No 4. P 14-15.
- Mouton, J. 2007. Approaches to programme evaluation research. *Journal of Public Administration*. Vol 42, No 6. pp. 490-511.
- Mouton, J. 2008. *Class and slide notes from the "Advanced Evaluation Course"* presented by the Evaluation Research Agency in Rondebosch, 20 – 24 October 2008.
- Nagel Stuart S & Teasley III C E. 1998. Diverse Perspectives for Public Policy Analysis, chpt 14 of Rabin Jack, Hildreth WB & Miller GJ, *Handbook of Public Administration*. 2nd edition. New York. Marcel Dekker. 507-533.
- Naidoo, I.A. 2007. *Unpublished research proposal, submitted to the Graduate School of Public and Development Management, University of Witwatersrand*.
- OECD. 2007. *OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes*. Paris, France: OECD.
- Owen. J.M. 2006. *Program Evaluation. Forms and Approaches. 3rd Edition*. New York: The Guilford Oress.
- Patton, M.Q. 2004. The Roots of Utilization-Focused Evaluation. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.

- Pierre, R.G. St. 2004. Using Randomized Experiments. In Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (eds) 2004. *Handbook of Practical Program Evaluation. Second Edition*. San Francisco: Jossey-Bass, John Wiley & Sons Inc.
- Posavac, E.J. & Carey, R.G. 1997. *Program evaluation. Methods and Case Studies*. 5th Edition. New Jersey: Prentice Hall.
- Reichardt, C.S. & Mark, M.M. 2004. Quasi-Experimentation. In Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (eds) 2004. *Handbook of Practical Program Evaluation. Second Edition*. San Francisco: Jossey-Bass, John Wiley & Sons Inc.
- Rogers, P.J. & Williams, B. 2006. Evaluation for Practice Improvement and Organizational Learning. In Shaw, I.F., Greene, J.C. and Mark, M.M. (eds). 2006. *The Sage Handbook of Evaluation*. London: Sage Publications.
- Rossi, P.H., Lipsey, M.W. & Freeman, H.E. 2004. *Evaluation. A Systematic Approach. Seventh Edition*. London: Sage Publications.
- Scriven, M. 2003. *Michael Scriven on the Difference between Evaluation and Social Science Research. The evaluation exchange*. Vol IX, No 4. p7.
- Segone, M. 2009. Enhancing evidence-based policy making through country-led monitoring and evaluation systems. In Segone, M. 2009. *Country-led monitoring and evaluation systems. Better evidence, better policies, better development results*. UNICEF Evaluation Working Papers. Romania: UNICEF.
- Shadish, W.R. (Jr), Cook, T.D. & Leviton, L.C. 1991. *Foundations of Program Evaluation. Theories of Practice*. California: Sage Publications.
- Stake, R. 2004. Stake and Responsive Evaluation. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Stufflebeam, D.L. & Shinkfield, A.J. 2007. *Evaluation Theory, Models & Applications*. San Francisco: Jossey-Bass.
- Stufflebeam, D.L. 2004. The 21st Century CIPP Model: Origins, Development and Use. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Valadez, J. & Bamberger, M. 1994. *Monitoring and Evaluating Social Programs in Developing Countries. A Handbook for Policymakers, Managers and Researchers*. EDI Development Studies. Washington: World Bank.
- Weiss, C.H. 1998. *Evaluation. Methods for studying programs and policies. 2nd edition*. New Jersey: Prentice Hall.

- Wholey, J.S. 2004. Using Evaluation to Improve Performance and Support Policy Decision Making. In Alkin, C.M. (ed). 2004. *Evaluation Roots, Tracing Theorists' Views and Influences*. California: Sage Publications.
- Wholey, J.S., Hatry, H.P. & Newcomer, K.E. (eds) 2004. *Handbook of Practical Program Evaluation. Second Edition*. San Francisco: Jossey-Bass, John Wiley & Sons Inc.