Chapter 2. Evaluation as science and as enterprise

2.1 Introduction

Although King Nebuchadnezzar practised Evaluation Studies millenniums ago in Babylon with Daniel's diet, even in our day management and entrepreneurs steer away from evaluation concepts and practices. According to Patton (1997:157) evaluation as science has much to offer, BUT the "shift of thinking to the evaluation terminology of outcomes and impact" proves difficult in enterprises with a long history of focusing only on profits, services, performance, activities and outputs.

Fortunately the traditional entrepreneurship environment is changing as "Evaluating AA" in the form of AccountAbility (AA) is being introduced with the establishment of AccountAbility Institutes and Corporate Citizenship Centres. Social responsibility, ethical standards, and environmentally sustainable development are the triple bottom line of non-financial accountability (AA 2002). This changing environment created a new demand for evaluations. According to Dollinger (1999:69), changes in business environment offer opportunities for entrepreneurs, although Chelimsky (1995:11) warns that changes in institutions "usually come slowly."

Evaluations should be mainstreamed in management and entrepreneurship, but are not easy as Sanders (2002:253) explains: "The practice of evaluation in organizations continues to be limited by perceptions that evaluation is a marginal activity. Arguments demonstrating the importance of evaluation have been ineffective in moving most organizations toward integrating evaluation into their daily routines. A multifaceted approach to making evaluation a part of organizational culture is proposed."

Chapter 2 deals with the following similar questions:

- WHAT are evaluation and development evaluation all about?
- HOW are evaluation, monitoring and surveillance being used?
- WHAT are valuation, accountability, responsiveness and triple bottom line?
- WHY is it necessary to be accountable and responsible?
- WHERE do evaluations take place? HOW are evaluators guided?
- FROM WHERE did evaluation originate?
- HOW, WHAT, and WHEN to evaluate? WHERE does evaluation fit in?
- HOW and WHERE should managers and entrepreneurs utilise evaluation?
- WHY are evaluators being misused? HOW can misevaluations take place?
- WHAT can be done to enhance the usefulness and influence of evaluation?

2.2 Defining evaluation and related concepts

According to De Vos (2000:365) evaluation is the "weighting or assessing the value of something." Patton (1997:23) defines program evaluation as the "systematic collection of information about the activities, characteristics, and outcomes of programs to make judgements about the program, improve program effectiveness, and inform decisions about future programming." Greene (2001:coverpage) refers to the evaluation practice as an "evaluation enterprise", thus an enterprise in itself. According to Stufflebeam (1994:323) 'Evaluation' should be narrowly and consistently defined. Patton, (2002a) elaborates on this by emphasising the common sense meaning of evaluation as the "worth of an object", and explains the origin of evaluation: "when somebody asked the question *Why*?" Early visions for evaluation focused on evaluation's expected role in "guiding funding decisions and differentiating the wheat from the chaff..." (Patton, 1997:12). Worthen (1997:5) agrees with them when defining evaluation: "to determine or fix value ... to examine and judge."

On a lighter note, for Halcolm, as denoted by Patton (1997:3), evaluation is a collective effort to outwit the following human propensities: "insidious prejudice, stultifying fear of the unknown, contagious avoidance, beguiling distortion of reality, awesomely selective perception, stupefying self-deception, profane rationalization, massive avoidance of truth ... all marvels of evolution's selection of the fittest."

Development evaluation in the World Bank (IPDET 2002:m1p3) is defined as: "A systematic search for answers about development interventions that involves gathering, analyzing, interpreting, and reporting information about quality." IPDET (2002 m3p51) defines Social Assessment as "a systematic assessment of the social processes and factors that could affect the outcomes of development projects."

Evaluations can also be evaluated. Evaluating evaluation itself is becoming popular. A meta-evaluation is defined by Patton (2002:211) as "an evaluation of an evaluation." Worthen (1997:450) agrees that no longer is meta-evaluation seen as merely "a nicety." It is now "an expectation." Utilization-focused program evaluation is defined by Patton (1997:23) as "evaluation done for and with specific, intended primary users for specific, intended uses."

IPDET (2002 m3p20) defines Rapid Assessments as "fairly quick and fairly clean" as opposed to "quick and dirty." Rapid Assessments are useful according to IPDET (2002:m3p16) because they intend to do evaluations quickly while obtaining

reasonably accurate and useful information, "by using a systematic strategy to obtain just essential information while the focus is on practical issues."

Not all endeavours are evaluable. IPDET (2002:m3p3) recommends that Evaluability Assessments be done and defines it as "a brief, preliminary study to determine whether an evaluation would be useful and feasible. An evaluability assessment identifies available data resources, clarifies key stakeholders' information needs, and considers the feasibility of different methods for conducting the evaluation."

For Bamberger (2002:1) "Shoestring evaluations" are strategies that can be used to reduce costs and time of an impact evaluation while ensuring acceptable levels of methodological rigor. A form of multiple-site program evaluation that originated in the W. K. Kellogg Foundation (WKKF) is a two-level approach that its developers termed cluster evaluation (Worthen 1997:474). Informal evaluation occurs whenever one chooses from among available alternatives without having somehow collected formal evidence about the relative merit of those alternatives (Worthen 1997:7).

Shadish (2002:1) defines evaluation even wider as "a profession composed of persons with varying interests, potentially encompassing but not limited to the evaluation of programs, products, personnel, policy, performance, proposals, technology, research, theory, and even of evaluation itself." A comprehensive evaluation is defined by (Baker 2000:1) as an evaluation that includes monitoring, process evaluation, cost benefit evaluation, and impact evaluation.

2.3 Evaluation, monitoring and surveillance

Evaluation as science also had its shifts as Patton (2002:151) revealed: "Indeed, in every arena ... international development-emphasis has shifted from providing services to attaining priority outcomes."

Schnoes et al (2000:96) noted that the concept 'evaluation' is used differently by different agencies and authors: "Some distinguish between 'monitoring' activities, which are conducted during project or program implementation to access the efficiency and effectiveness with which inputs are used to achieve intended outputs, and 'evaluation' activities, which assess the extent to which projects or programs have achieved their intended objectives and have produced their intended changes and benefits in the target populations. In other cases the term evaluation is used more broadly to cover both of these functions." For Baker (2000:1) a program monitoring system enables continuous feedback on the status of program

implementation, identifying specific problems as they arise, while process evaluation is concerned with how the program operates and focuses on service delivery.

For OED (2002b:5) Monitoring and Evaluation of development activities provide government officials, development managers, and civil society with better means for learning from past experience, improving service delivery, planning and allocating resources, and demonstrating results "as part of accountability to key stakeholders."

The Development Bank of Southern Africa (OEU, 2000) defined the following:

- MONITORING is a continuous assessment of activities especially done during the Activity and Output phases. It is an essential management tool and part of good management practice and day-to-day management. Monitoring will provide the basis for corrective actions, mainly during construction.
- SURVEILLANCE is similar to monitoring and means 'to observe' but done after completion, during the lifespan of the project or loan, usually 20 years.
- EVALUATION is an assessment of an ongoing or completed project against stated project objectives, the project goal, and the performance indicators contained in the project description and the loan agreement. Evaluations are done after the project was allowed enough time to make an impact.

Participatory Monitoring and Evaluation (PME) is defined by IPDET (2002:m12p89) as: "A broad constellation of approaches, methods and techniques that can be used to strengthen poverty alleviation programs, ensure accountability, build local capacity, and foster an environment of partnership and collaborative learning" as "Most of the time, the outside experts fly in and out without taking the time to listen to the people."

Odwedo (2000:81) agrees with the above definitions: "Monitoring and Evaluation (M&E) is a process of assessing a project or policy and taking any corrective action required. It involves collecting and analysing information about a project and generating recommendations for change. Monitoring is usually conducted as an ongoing activity, throughout the life of the project, whereas evaluations are undertaken at certain intervals such as project midterm and completion."

2.4 Valuation versus evaluation

The concept of value is more familiar to business management and entrepreneurs than the concept evaluation. For De Vos (2000:365) the two are almost similar as "evaluation is the weighting or assessing the value of something." Valuation refers to the estimation of a thing's worth, or prize set on something. According to Todaro

(2000:8) developing economics must be mindful of the crucial roles that "values, attitudes and institutions" play in the overall development process. Values, principles, standards, or qualities are considered worthwhile and desirable. For Wickham (2001:9), an opportunity is a gap in a market where the potential exists to do something better and create value. A value judgement reflects personal or class beliefs in normative economics, according to Todaro (2000:769). According to Kuratko (2001:157) "Successful entrepreneurs, whatever their individual motivation - be it money, power, curiosity, or the desire for fame and recognition, try to create value and make a contribution." Makhubela (2001:1) indicates that as the knowledge economy forces a radical rethinking of organisational value (inclusion of intangible assets and resources), "there is realisation that an organisation's value consists of more than what is shown in its traditional balance – and value sheets (only the tangible assets)."

Kim & Mauborgne (1996:106) distinguish between Conventional Logic versus Value Innovation. "Conventional Logic leads companies to compete at the margin for incremental share. The logic value innovation starts with an ambition to dominate the market by offering a tremendous leap in value. Value innovators never say, here's what competitors are doing; let's do this in response. They monitor competitors but do not use them as benchmarks."

Wickham (2001:6) defines the *Entrepreneurial process* in which the entrepreneur engages, as "the means through which new value is created..." Hisrich (1998:9) agrees with this by saying that "*Entrepreneurship* is the process of creating something new with value", by devoting the necessary time and effort, assuming the accompanying financial, psychic, and social risks, and receiving the resulting rewards of monetary and personal satisfaction and independence.

Businesses, business managers and entrepreneurs add value and their endeavours can thus be evaluated. Evaluators also add value to future endeavours and projects by evaluating and assessing the outcomes of existing endeavours and projects.

2.5 Evaluation versus accountability, responsiveness and triple bottom line

The traditional entrepreneurship environment is changing with new kind of "Evaluating AA" in the form of AccountAbility (AA). The three issues: social responsibility, ethical standards, and environmentally sustainable development are the triple bottom line of non-financial accountability (AA 2002).

According to Shay (2001:2) social accounting should be regarded as a "practice of excellence" in business. But according to Zadek (2002:1) there are real concerns as to the future course of corporate responsiveness and responsibility as a means of addressing major social and environmental challenges. Even most inveterate inside players were vocal in challenging the adequacy of progress made to date. Such challenges are undoubtedly good news for corporate evaluators.

Todaro (2000:710) stresses the fact that "managers must be made more accountable for resource allocation and investment decisions." For Mayne (1999:9) "Accountability for results is an essential feature of managerialism." On the other hand evaluation serves to identify strengths and weaknesses, highlight the good, and expose the faulty, but cannot single-handedly correct problems, for that is the role of management and other stakeholders according to Worthen (1997:23).

According to Longenecker (2003:524) the first element of AccountAbility, social responsibility, reconciles the need for profit with social obligations, protect interests of customers, employees, suppliers and general public and includes environmental protection and pollution aspects. Managers need to meet regulations even if they are costly, because mistakes can be more costly in terms of legal claims. Socially acceptable actions create goodwill in the community and attract customers, they comply with government, industry and other regulations, are responses to community needs and contribute to community organisations.

Traditionally, accountability has been viewed as "something negative done" to someone, according to Mayne (1999:159). People were held accountable and blame is melted out for failures. Accountability was seen as control and in this view had a rather negative force, something any sensible manager seeks to avoid if possible, as this traditional accountability, besides being an annoyance, is probably seen as of secondary importance to those managers trying to motivate people in accomplishing objectives.

Most people would agree that some amount of regularity is needed, but are quite content to leave it for auditors to handle. Mayne (1999:159) argue strongly against the auditing approach because a mindset of holding all staff accountable for correct procedures appears anti-ethical. "The most effective way to hold employees accountable is to make them feel accountable." They want to be accountable because it is the only way for them, as for us all, to be important (Mayne 1999:160). Accountability is closely related to ethical behaviour and integrity.

2.6 Ethical versus legal concepts in evaluation

For Longenecker (2003:524) ethical issues are questions of right or wrong, questions of integrity, and are going beyond what is legal or illegal. IPDET (2002:m12p38) notes "Ethics are complicated...No single law regulates ethical behavior of evaluators...Behavior can be legal but unethical." Longenecker (2003:526) states four dimensions of integrity: being honest; being truthful; being respectful; and being fair.

"Taking an ethical stand in today's materialistic world can be very costly. Two out of every three whistle-blowers in the past lost their jobs in the organisations whose wrongdoing they exposed" alert Lumsdaine & Lumsdaine (1995:329).

Worthen (1997:291) refers to Morris and Cohn's (1993) survey of AEA (American Evaluation Association) members that netted 459 responses. Of those respondents, nearly two-thirds reported they had encountered ethical problems in their evaluation work. Many of these problems reflect unethical conduct by evaluation participants other than the evaluator. This study found for example:

- "Evaluator is pressured by stakeholders to alter presentation of findings."
- "Findings are suppressed or ignored by stakeholder."
- "Findings are misused by stakeholder."
- "Findings are used to punish evaluator."
- "Findings are used to punish someone other than evaluator."
- "Findings are deliberately modified by stakeholder prior to release."
- "Findings are misinterpreted by stakeholder."
- "Stakeholder misrepresents authorship or plagiarizes report content."
- "Stakeholder prejudges what findings "should be."
- "Stakeholder prejudges what findings in an ethically questionable fashion."
- "Stakeholder declares certain evaluative questions "off-limits", despite their obvious relevance."
- "Sponsors omit other legitimate stakeholders from planning process."
- "Stakeholder pressurizes evaluator to violate confidentiality" (Worthen, 1997).

It is difficult for an evaluator to survive in such a hostile environment, especially when the evaluation community is expecting innovative approaches in evaluations. Russell (1999) notes that innovation is associated with better financial performance in dynamic environments, but is not associated with increased performance in hostile or static environments.

For Eloff (2001:6) the new economy presents challenging opportunities for countries, big business, small businesses and entrepreneurs, but it also requires more from companies "in terms of moral imperatives." AccountAbility (2001b) agrees that trust is an essential ingredient for positive interaction between an organisation and all the people it affects by its activities.

Longenecker (2003:524) states some general principles in dealing with ethics:

- Consider welfare of those around you and do what is right.
- How would you feel if your decisions were to be published in the Daily News?
- How would you explain your acts or decisions to your mother?
- Do the right thing; stick to your principles; principles are not for sale.

According to House (2000:11) the evaluator is "not a passive bystander, an innocent facilitator, or a philosopher king who makes decisions for others, but rather a conscientious professional who adheres to a set of defensible, carefully considered principles for enhancing inclusion, dialogue and deliberation." According to Worthen (1997:327) it is appropriate to suggest that the ultimate ethical principal, "Do unto others as you would have them done to you" is more or less binding on evaluators.

In accordance with IPDET (2002:m12p28) "Ethics always represent a choice" and "Ethical standards provide guidelines for making those choices." For these reasons standards and guiding principles for evaluations are of fundamental importance.

2.7 Standards and guiding principles for evaluations

Evaluators cannot make or break as they wish. The International Business Leaders Forum, IBLF (2002), often discussed good corporate governance and exchange responsible evaluation and business practices. For Bastoe (2000:117) good evaluations constitute: ethics, quality standards, use and dissemination.

According to Shadish (2002:1) the purpose of documenting guiding principles is "to foster continuing development of the profession of evaluation, and the socialization of its members. The principles are meant to stimulate discussion and to provide a language for dialogue about the proper practice and application of evaluation among members of the profession, sponsors of evaluation, and others interested in evaluation."

The following Standards and guiding principles for the evaluation enterprise that was developed by the American Evaluation Association (AEA, 2002b) can be useful for any enterprise:

- Systematic Inquiry: Evaluators conduct systematic, database inquiries about whatever is being evaluated.
- Competence: Evaluators provide competent performance to stakeholders.
- Integrity/Honesty: Evaluators ensure the honesty and integrity of the entire evaluation process.
- Respect for People: Evaluators respect the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders.
- Responsibilities for General and Public Welfare: Evaluators articulate and take into account the diversity of interests and values.

The American Evaluation Association (AEA) Task Force on Guiding Principles for Evaluators (Shadish 2002:1) investigated the fact that "these principles were developed in the context of Western cultures, particularly the United States, and so may reflect the experiences of that context." The relevance of these principles may vary across other cultures, other continents and even across sub-cultures within the United States.

Bastoe (2000:120) mentioned that the OECD public sector management group (PUMA) developed a set of guidelines called *Best practice guidelines for evaluation*. "PUMA sees evaluation as integrated in a results-oriented environment because it provides feedback on the efficiency, effectiveness and performance of public policies and can be critical to policy improvement and innovation. It contributes to accountable governance."

Other evaluation societies have been through long processes to develop their evaluation standards. The Canadian Evaluation Society has developed what they call "guidelines for ethical conduct" (Bastoe 2000:119), including three concerns:

- Competence; evaluators are to be competent in their provision of services.
- Integrity; evaluators are to act with integrity in their relationships.
- Accountability; evaluators are to be accountable for their performance and their product (AA 2002; Bastoe 2000:117).

A review of the relevance of the Programme Evaluation Standards (PES) to evaluation work in Africa was undertaken in a workshop setting, at the Inaugural Conference of the African Evaluation Association and in several meetings of the Kenya Evaluation Association. Most of the PES was accepted as currently specified. Modifications of some standards were proposed by Patel & Russon (2000:125).

2.8 Evaluation as inspection versus evaluation as research

According to (Worthen 1997:27) in the public sector, formal evaluation was evident as early as 2000 B.C., when Chinese officials conducted civil service inspections and examinations to measure proficiency of public officials, and in education, Socrates used verbally mediated evaluations as part of the learning process. But centuries passed before formal evaluations began to compete with religious and political beliefs as the driving force behind social and educational decisions (Worthen 1997: 27).

It is important for managers and entrepreneurs to know the background of evaluators as well as the ideologies of specific development finance institutions or international funding institutions. There are historical reasons why they place emphasis on different issues. Evaluation science as it is today developed from two main schools:

- The American School: Evaluation became mutations from research. Evaluations were judged on their methodology and research background.
- The English School: Evaluations were based on the school inspection system of smooth operation, auditing and control (Patton, 2002a).

Some evaluation associations have their roots in research. According to Shadish (2002:1) the Evaluation Network (ENet) and the Evaluation Research Society (ERS) merged in 1986 to create the American Evaluation Association (AEA).

Evaluation is a profession "composed of persons with varying interests", potentially encompassing but not limited to the evaluation of programs, products, personnel, policy, performance, proposals, technology, research, theory, and even of evaluation itself (AEA, 2002b:2). The intellectual roots of evaluation and the prominent arena of the evaluation discipline, are traced to the 1960s, the "era of the Great Society, and to what the evaluation discipline has referred to as ... Experimenting Society", according to Caracelli (2000:99).

Worthen (1997: 28) argues that the late 1800s also saw the beginning of efforts to accredit U.S. universities and secondary schools, although accreditation did not really become a potent force for evaluating educational institutions until several strong regional accrediting associations were established in the 1930s.

On a lighter note Worthen (1997: 25) observes: “Evaluation, as an established field, is now in its late adolescent years. The bubbling, exiting, fast developing childhood years of the late 1960s and early 1970s gave way in the mid to late 1970s to the less assured, serious, introspective early adolescent years.”

Nowadays evaluation is according to Patton (1997:103) observed by its "emphasis on reality testing based on systematic data collection for improvement, judging merit and worth, or generating knowledge about effectiveness." The processes of evaluation support change in organizations by getting people engaged in "reality testing", that is, helping them think empirically, with attention to specificity and clarity, and teaching them the methods and utility of data-based decision-making.

2.9 Types of evaluations: Cluster, multi-site and sectoral (theme) evaluations

Multi-site Evaluations are for IPDET (2002:3-6) important because rather than look at a single intervention, it is sometimes more useful to look at interventions that have been implemented in a variety of locations. The intervention may have been implemented in the same way in all locations or implemented slightly differently in each location. "This type of evaluation provides information about the overall experience of the intervention as well as a deeper understanding about the variations."

Cluster evaluations are similar to multi-site evaluations but the intention is different. Like multi-site evaluations, cluster evaluations focus on interventions that share a common mission, strategy and target population. However, the evaluation is not intended to determine whether an intervention works or to assure accountability. It does not evaluate the success or failure of individual interventions nor does it identify interventions to be terminated. Its intent is to learn about what happened across the clusters and to ascertain lessons learned. Information is only reported in aggregate so that no one project is identified (IPDET 2002:3-6). Cluster evaluations originated in the W. K. Kellogg Foundation (WKKF) as a two-level approach (Worthen 1997:474).

IPDET (2002:3-6) classifies Sector or Thematic Evaluations as studies that can compare experiences across countries about a sector or theme such as health, nutrition and population study.

The evaluation process also reveals different types of evaluations.

2.10 The Evaluation Process: How, what, and when to evaluate

The Evaluation Process is similar to the following Research Process described by Coopers & Schindler (2001:61) as:

- Management Dilemma;
- Research Questions;
- Management Questions;

- Investigative Questions;
- Measurement Questions; and
- Management Decision.

How? According to IPDET (2002:m2p5) the Evaluation Process consists of the following:

1. Planning the Design:
   - Understand context, develop logic model;
   - Assess stakeholders' needs;
   - Identify evaluation questions;
   - Select appropriate design, measures, criteria;
   - Develop data collection strategy, including sampling plan;
   - Develop data analysis strategy; and
   - Prepare work plan.
2. Doing:
   - Gather the data;
   - Prepare data for analysis;
   - Analyse data;
   - Interpret the data; and
   - Formulate findings.
3. Reporting the results:
   - Major findings: what works, what doesn't;
   - Clear, simple language;
   - Use of charts and tables to highlight major findings; and
   - Plan for dissemination.
4. Recommendations:
   - Clear and specific;
   - Who should do what;
   - Evidence to support recommendations; and
   - Logical relationship between recommendations.
5. Feedback process to:
   - Stakeholders;
   - Project managers,
   - Donors, Officials,
   - Beneficiaries, members of community.

What to Evaluate? According to IPDET (2002:m2p5) the following aspects are candidates for evaluation

- A single intervention or project in one location or a single project implemented in several locations.
- Intervention and programs comprised of various activities which are intended to contribute to a common goal.
- Organizations and multiple intervention programs delivered by an organization.
- Themes and sectors; evaluations of interventions across a specific policy arena, such as education, forestry, agriculture, and health.
- Country assistance; evaluations of progress relative to the plan, the overall effect of aid, and Lessons learned.

When to Evaluate? According to IPDET (2002:m2p5) the best time to evaluate is:

- Before development intervention starts to improve design;
- During the implementation to improve implementation and to identify barriers to be removed;
- Mid-term evaluation to determine relevance, effectiveness, efficiency, lessons learned; as a management tool;
- Terminal evaluation at the end of intervention to determine relevance, effectiveness, efficiency, early signs of impact and sustainability and to obtain lessons learned for future projects;
- Ex-post evaluation is conducted two or more years after the completion of mature interventions and is preferred for clusters, geographical location or theme and judges relevance, performance and success, lessons learned for future policy and for formulation or programming.

According to Bamberger (2002:6) in many cases good estimates on most or all of the evaluation questions can be obtained with relatively simple evaluation designs. Obviously the larger and more complex the project, the longer the time period being studied and the more diverse the areas in which it is operating, the more important it becomes to use more rigorous evaluation designs.

2.11 Evaluation as enterprise; developing a niche and scope

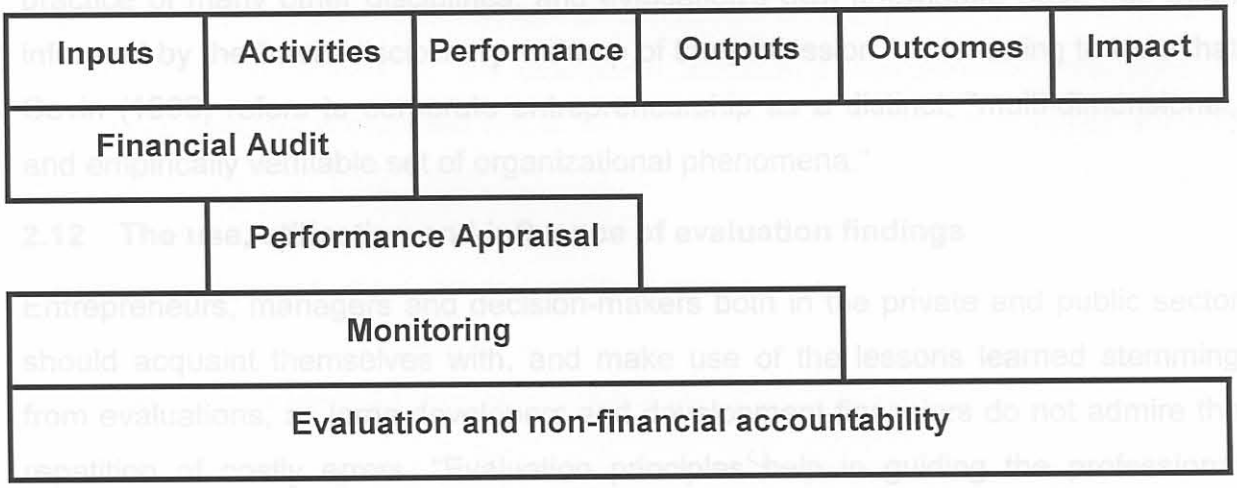
An evaluator can also be an entrepreneur. For Hisrich (1998:9) entrepreneurs are found in all professions - education, medicine, research, law, architecture, engineering, social work, and distribution, and therefore also in evaluation. Evaluation

entrepreneurs are in dire need for development success because the failure of many development efforts has prompted the development community to search for a "new paradigm to guide the formulation, implementation and *evaluation* of development strategies" (Essama-Nssah 2000).

Evaluation should span over the whole spectrum of an operation, much wider than audit. Auditors across the world are beginning to include evaluation – which they call "performance audit," or "value for money audit" or "comprehensive audit" – as one of their key work areas... "and auditors are recognizing more and more that an evaluation is not at all the same thing as an audit" (Chelimsky, 1995:3).

Auditing is according to IPDET (2002:m1p4) rooted in financial accounting concepts such as compliance, verification, internal controls and good management practices. It compares "what is" to "what should be." Performance audit is an objective and systematic examination of evidence to ensure accountability.

Figure 2.1: The niche and scope of the evaluation practice



(Developed from the World Bank OED model: Picciotto & Rist, 1995:240).

Evaluation differs from auditing in that it is "not necessarily based on known criteria", but starts with clarification of criteria. Bastoe (2000:120) defines auditing as "essentially normative, examining the match or discrepancy between the criterion and the actual condition."

Based on differences in training, experience, and work settings, the profession of evaluation encompasses diverse perceptions about the primary purpose of evaluation. According to Shadish (2002:1) "These include but are not limited to the following: bettering products, personnel, programs, organizations, governments, consumers and the public interest; contributing to informed decision-making and

more enlightened change; precipitating needed change; empowering all stakeholders by collecting data from them and engaging them in the evaluation process; and experiencing the excitement of new insights. Despite that diversity, the common ground is that evaluators aspire to construct and provide the best possible information that might bear on the value of whatever is being evaluated." Lacking a single intellectual rallying point, development professionals must inevitably trespass across disciplines as "no single discipline can claim to dominate an endeavour that deals with the multiple challenges, hopes, and exertions of most of human-kind" (Picciotto & Rist, 1995:x).

For Greene, (2001) the "evaluation enterprise" is becoming "big business", and evaluators are like entrepreneurs and corporate entrepreneurs, also looking for opportunities.

Evaluation is not a discipline on its own. Caracelli (2000:103) describes evaluation's niche as "transdiscipline." Evaluation is actually interdisciplinary as it informs the practice of many other disciplines, and evaluation's own knowledge base has been informed by the "multi-disciplinary makeup of the profession." Interesting to note that Covin (1999) refers to corporate entrepreneurship as a distinct, "multi-dimensional, and empirically verifiable set of organizational phenomena."

2.12 The use, utilisation and influence of evaluation findings

Entrepreneurs, managers and decision-makers both in the private and public sector should acquaint themselves with, and make use of the lessons learned stemming from evaluations, as large developers and development financiers do not admire the repetition of costly errors. "Evaluation principles help in guiding the professional practice of evaluators, inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators" (Bastoe 2000:118).

The reason for paying so much attention to user's milieu is that "policy makers and evaluators are quite different with respect to their goals" (Chelimsky, 1995:6). To the user, evidence based on a thorough evaluation may merely be instrumental to a negotiation or a decision, while to the evaluator's evidence it is the end in itself. Evaluators do not invoke certainty unless their data provide such certainty, but decision- and policy makers may place positive value on making a decision, "regardless of whether there is strong objective evidence to support that decision"

(Chelimsky, 1995:6). Another possible reason why evaluation is not frequently used, is the "difficulty of the task" (Squire 1995:47).

To be useful, evaluation must add something to the process or function to which it is linked (Picciotto & Rist, 1995:x). Like 'development', 'evaluation' is used in several dimensions. Historically, conversations about influence have occurred under several themes: "internal and external evaluation, evaluators roles, evaluation as a profession, ethics and values, and use of results" (Kirkhart 2000:5). The domain of development hinges on the notion of the things that are worth promoting (Essama-Nssah 2000).

"Evaluators are trained, are least in part, to be sceptics, to raise insightful (one hopes) questions that otherwise may never have been considered" according to Worthen (1997:255). This training is never more valuable than during the divergent phase of identifying evaluation questions and criteria, for some important questions may be omitted unless the evaluator raises them himself.

A problem hampering the utilisation of evaluation in many countries is a lack of interest and commitment to the evaluation function at the political level (Bastoe 2000:117). The space engineer Roger Biosjoly's evaluation was timely but in vain. He desperately tried to stop Challenger Space Shuttle's launch when he discovered problems, but his monitoring and evaluation of the fatal problem was ignored. The Challenger space shuttle exploded while launched in January 1986. Roger "left the space programme and became a champion for training professionals in ethical sensitivity" (Lumsdaine & Lumsdaine 1995:239).

According to IPDET (2002:m2p5) the following institutions use and should benefit from evaluation: Government officials; Parliament; Program managers and staff; Citizens; Businesses; NGOs; Civil society; Donors; Participants.

For Patton (1997:79) evaluation findings have mainly three uses: To render judgement, to improve programs, and to generate knowledge-support decision-making (enlightenment).

- Rendering judgement: This is about overall merit or worth of the program or endeavour. It might influence a major decision whether it should be continued, enlarged, disseminated, or terminated.

- Improvement: Decisions about how to improve a program tend to be made in small, incremental steps based on specific evaluation findings aimed at instrumental use.
- Enlightenment: Policy decisions informed by cumulative knowledge from evaluation imply a weak and diffuse connection between specific evaluation findings and the eventual decision made.

Using monitoring and evaluation findings on similar projects are extremely important in any entrepreneurial endeavour. It is simply too costly to ignore the lessons learned and to make the same mistakes.

Worthen (1997:23) warns that the so-called usefulness of evaluation has led some persons to look to it as a panacea for all the ills of society. "But evaluation alone cannot solve all the problems of society. One of the biggest mistakes of evaluators is to promise results that cannot possibly be attained."

2.13 Misusing and harassment of evaluators for misrepresentation

"Misuse and harassment of evaluators still happen everywhere" (Bastoe 2000:117). Organisations and management want to hear good things about themselves. Bad news is not appreciated and often sidelined. The Internal Recognition and Reward (IRR) and performance systems of corporations place much emphasis on positive things, therefore evaluation processes and findings can be misused in the search for political and economic advantage.

According to Worthen (1997:230) the political nature of decisions should not be underestimated: E.g.: Whose values are attended to, how they are weighted, what variables are studied, how information is reported and to whom, how clients and other audiences intend to use evaluative information, what kind of support is given to the evaluation and by whom, what potentially embarrassing information is hidden, what possible actions might be taken to subvert the evaluation, and how the evaluator might be co-opted by individuals or groups? (Worthen 1997:230).

Lumsdaine & Lumsdaine (1995:329) warned that unfortunately evaluators are blackmailed in organisations that do not like their ethical standpoints and many of these "ethical resisters" were fired in the past without the prospect of ever getting a job again!

The basic obstacles to a more influential evaluation function are institutional constraints and with the frequent lack of "effective communications between

evaluators and decision makers." Ethical standards are of crucial importance to evaluators. They need to "ask the right questions, avoid emotionally and politically charged issues, use the right mix of evaluation methods, and make lucid, credible presentations" (Picciotto & Rist, 1995:x).

According to Sandstrom (1995:14) managers and critics can sometimes use the information in an unfair or distorted manner. Management often misuse evaluators to paint a colourful picture of enterprise endeavours. Evaluation findings, often distorted, are main headings in many annual reports. Being more open and rigorous about evaluation does not always make life comfortable, especially where evaluators need to question the endeavours of colleagues.

Evaluators have to make use of all kinds of participatory methods to get their message across to top management and board, without being fired by middle management. It is important to take note that the role of the evaluator is expanding from "dispassionate outsider to co-investigator with program staff, who assumes a variety of roles that require degrees of engagement with stakeholders" (Caracelli 2000:103).

Zadek (2002:2) warns that like all movements in their stages, "the field of corporate responsibility has been dogged by over-idealism, evangelism, and a lack of serious self-critique." Good communication is necessary to overcome the negative ideas on evaluation. Russell (1999) states that communication and scanning are positively correlated with corporate entrepreneuring while control is negatively correlated.

It is not easy for evaluators and managers to trust one another. "I would still advise any evaluator to remember that when the lion and calf lie down together, the calf isn't likely to get much sleep" warns Chelimsky (1995:11).

This is why Evaluation Associations set "Propriety Standards" to curb misuse. The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results (AEA, 2002a:3).

Evaluators have become self-consciously aware that evaluation studies are often misused or ignored, with the result that some individuals have argued for decreased emphasis on the evaluative process. "But that seems no more sensible than abandoning medical diagnosis because science has not yet successfully eliminated all disease" (Worthen 1997: 510).

2.14 Misevaluations and misunderstanding evaluations and evaluators

Alkin (1988) referred to by Patton (1997:359-360) made a critical distinction of *misevaluation*, in which an evaluator performs poorly or fails to adhere to standards and principles, and *misuse*, in which users manipulate the evaluation in ways that distort the findings or corrupt the inquiry. Misevaluation includes the oversight and omission of essential information that might guide the evaluation findings in another direction. This often happen when the evaluators or team members are not independent and are prescribed by managers who obviously have hidden agendas. In the past problems with results-based use of evaluation results caused some mistrusting of the evaluator (Caracelli 2000:103). The performance based recognition and rewards systems are biased towards institutional upholding and do not tolerate negative criticism on endeavours. According to Todaro (2000:8) "Institutional and structural problems and the power of historical, cultural and religious forces should not be underestimated."

Misunderstandings and misperceptions often result between evaluators on the one side and managers, entrepreneurs and decision-makers on the other side, which lead to misevaluations. "Yet unless policymakers, managers, and staff, both in developing countries and development agencies, internalise the processes and lessons of evaluation, evaluation will have limited benefits on the ground" (Picciotto & Rist, 1995: xiii-xiv).

Reliable evaluation results will have to become embedded within the values, processes, and incentive framework that inform the decision making process. Evaluators need to write reports in such a way that everybody will feel good. Especially in an African context, Patel & Russon (2000:125) suggest that an evaluative report might be "selectively supportive, rather than critically comprehensive."

For Cronbach (1982:239), a pioneer on measurement, evaluation is as much art as science: "Developing an evaluation is an exercise of the dramatic imagination." Data and indicators can misrepresent. Indicators should be accurate because it presents a balanced presentation of financial, operational and impact performance, "what you measure is what you get", summarise complex information "at a glance" and ensures evaluability of projects (ITAD, 1999).

Validity of evaluation findings is important. Bamberger (2002:11) warns that "In their efforts to reduce time and costs evaluators have frequently ignored some of the basic principles of evaluation design such as: random sampling, specification of the evaluation model, instrument development, and full documentation of the data collection and analysis process. As a consequence many rapid evaluations suffer from serious methodological weaknesses which threaten the validity or generalizability of evaluation findings."

Patton (1997:251) refers to Sir Josiah Stamp an English economist who said already in 1911: "The government ministries are very keen on amassing statistics. They collect them, raise them to the n-th power, take the cube root, and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first place from the village watchman, who just puts down what he damn well pleases."

Alkin (Patton 1997:251) studied utilisation and found that "for evaluations to have impact, users must believe what evaluators have to say." The "believability" of an evaluation depends on much more than the perceived scientific validity of the data and findings. Believability depends on the users' perceptions of and experiences with the program being evaluated, users' prior knowledge and prejudices, the perceived adequacy of evaluation procedures, and the users' trust in the evaluator. Trust, believability, and credibility are the "underpinnings of overall evaluation validity."

Without wide participation by members of all partner organisations in the processes of planning the evaluation, the principles and priorities of each party will not be fully reflected in the final plan. Even the Logframe can "conceal differences in interest of the various actors involved, hence, dangerously ignoring potential areas of conflict that may jeopardise the success of the intervention" (Pasteur 2001).

2.15 Enhancing the usefulness and influence of evaluation

Evaluations should be used and be useful for managers, entrepreneurs and decision-makers. For Patton (1997:20) utilization-focused evaluation begins with the premise that evaluations should be "judged by utility and actual use", therefore, "evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will affect use."

Creating a successful evaluation function means "bringing together skilled evaluators and helping them understand each other; choosing topics to evaluate that are both important in a policy sense and evaluable; finding the time, funding, and user

commitment to do the job properly; protecting the independence of the function; supporting evaluators in fighting against distortion of their findings; and ensuring that evaluators understand the culture and information needs of sometimes far-away users, on whom the impact of their findings may depend" (Chelimsky, 1995:11). Simple methods and graphical presentation of findings may be needed to communicate best (Scheifer 2000:143).

The construct of 'use' itself has multiple attributes (Caracelli 2000:102). Kirkhart's (2000) integrated theory of influence represents an incorporation of past understandings of use based on evaluation literature, as well as providing a deeper, more developed and differentiated construct of use in light of the changing circumstances of the field and its increasing diversity, argues Caracelli (2000:102).

Useful evaluations as well as "Evaluation Impact" are important for the America Evaluation Association. "Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased" (AEA, 2002a). Since the objective is to empower the poor, their views should play an important role in the evaluation (Squire 1995:47). It is suggested by Scheifer (2000:139) that performance measure data that are already being collected could become much more useful, by further analysing the relationships among several types of measures, and then analysing variability among program delivery units for those measures by using logic models (Please refer to Chapters 6 & 7).

Useful evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood (AEA, 2002a). Beneficiary assessments could help to identify projects for which the assessments are likely to be most valuable and could measure the return (Squire 1995:51). Patton (1997) describes Stake's responsive evaluation (1975) as "an alternative to the dominant experimental paradigm and one that influenced evaluators to think about the connection between methods and use" (Caracelli 2000;101).

According to IPDET (2002:m12p22) the use of evaluation can be improved by "gaining support from the top, increase their awareness of the role, evaluations can play and how it can help them and help them set realistic expectations." Evaluations should be on time according to IPDET (2002:12-25) because "A good evaluation that arrives after the decision has been made is useless."

For evaluation findings to be used more in policy making, it is "almost as important for evaluators to understand their user – and thereby be able to explain their findings in terms that make sense to him or her – as it is for the findings to be methodologically strong and compelling in their own right" (Chelimsky, 1995:8).

To enhance evaluation influence, it should be conceptualised along three dimensions (Caracelli 2000:102):

- Source: addressing results-based and process-based influence;
- Intention: addressing intended and unintended influence; and
- Time: addressing influence that occurs during evaluation, at the end of evaluation, and in the future.

The Evaluator should according to IPDET (2002:m12p98) be a facilitator that:

- Is self-aware and self-critical;
- Promotes mistakes as opportunities to learn;
- Focuses on process, not outcome; and
- Sits, listens, and learns.

To maximize the usefulness of evaluations stakeholders at the "top, bottom, and side" should be involved according to IPDET (2002:m12p22). According to Caracelli (2000:102) the shift in terminology from use to influence creates a broader framework that allows for multiple perspectives.

Corporations that use corporate evaluators are in many cases progressive institutions that also know the benefits of corporate entrepreneurs to optimise on innovation, creative ideas, effectiveness and efficiency.

2.16 Similarities between the corporate entrepreneur and the corporate evaluator

Evaluators can learn from management and entrepreneurial constructs. The corporate entrepreneur (intrapreneur) as construct is important to discuss due to the similarities between the problems and experience of the evaluation construct.

Although the constructs of entrepreneurship has been limited to new venture creation by some scholars, according to Dess (1999), Corporate Entrepreneurship (CE) may be viewed more broadly as consisting of the birth of new businesses within existing organizations and the transformation of organizations through strategic renewal, i.e., the creation of new wealth through the combination of resources. Barrett et al (2000) sees CE wider: Researchers have defined numerous dimensions of individual

entrepreneurship that translate well to corporate entrepreneurship. CE can be defined as an organizational process that encourages and practices the utilization of innovation, constructive risk-taking, and pursuit of new opportunities.

Russell (1999) uses a cognitive mapping approach to build a model of corporate entrepreneurship from an organizational perspective. This approach analyzes corporate entrepreneuring from a systems viewpoint and facilitates an understanding of the process through which entrepreneurial firms generate innovation. Covin (1999) agrees with Russell (1999) in defining corporate entrepreneurship as the presence of innovation, and adds competitive superiority to his definition. "Fostering intrapreneurial behaviors and practices has assumed prime importance in the grand strategies of many firms where creating innovation is perceived as an important means of establishing and maintaining competitive advantage as well as a method for initiating corporate renewal" argues Russell (1999). Covin (1999) argues that innovation, broadly defined, is the single common theme underlying all forms of corporate entrepreneurship.

Schindehutte et al (2000) state that entrepreneurship can be found in any established organization, and not only businesses. She stresses the concept of "entrepreneurial intensity", which focuses on both the frequency and degree of entrepreneurship in a company or organisation.

The role of the evaluator, especially the corporate evaluator, fits neatly into the definitions of a corporate entrepreneur. Evaluations should be innovative; it should create new ideas on old issues, and should investigate alternative ways to reach effectiveness. Although difficult to achieve, obtaining the "appropriate level of venture autonomy can reap great rewards" according to Simon (1999).

Another similarity between intrapreneurship and evaluation is sometimes the "corporation's unwarranted prosecution" (Simon 1999), the impatience, jealousy and suspicion of middle management, the lack of adequate funds, and what Worthen (1997:291) calls punishment, misinterpretation, prejudgement and hostile environments that the evaluator must endure. Simon (1999) suggests a managerial approach that will stimulate innovation and "re-energise employees."

To counteract the above negative aspects, Simon (1999) suggests three roles to protect the corporate entrepreneur (venture management) as well as the company. Firstly, the Management will be responsible for the funding and general management.

Secondly, the Godparent will be an older and senior staff member that will act as lawyer for the intrapreneur to ensure that the plug will not be pulled too soon on the venture, and thirdly, the Ombudsperson that will act as a judge between the Management and the Godparent. According to Simon (1999) the ombudsperson should manage the tension between the godparent and corporation to provide venture management the necessary freedom and support to succeed and corporate management the necessary controls to avoid large losses.

Evaluators doing innovative work in hostile environments can benefit from a Godparent and Ombudsperson taking care of their ventures!

2.17 Conclusions

Adopting evaluation concepts is not an easy process as Worthen (1997: 510) agrees that "systems have most of the earmarks of classical bureaucracies and, historically, have been reasonably successful in resisting change in practices and policies." Recently strong social forces have coalesced to push many systems out from behind their barriers; change has become a much more frequent reality. However, without a tradition of planned programs, the changes that are occurring can be often little more than random adoption of faddish innovations. Perhaps the most important deficiency, which fosters such a situation, is the lack of dependable information in the performance of available products, practices, and programs. Without such information, practitioners cannot readily correct deficiencies in present pricks or intelligently select new products or practices for adoption.

Evaluation holds great promise in providing stakeholders with badly needed information, which can be used to improve the processes of human service management and entrepreneurship. This is of utmost importance, although not an easy task as Sanders (2002:253) explained. While obviously not a panacea, in line with Worthen (1997: 510), he agrees that evaluation can have a profound impact on the human services professions, including management and entrepreneurship.

Chapter 2 intended to present a theoretical foundation of evaluation as science but also as enterprise in its own right. Chapters 3 and 4 will investigate the place and propagate the extensive use of evaluation concepts in development, management, entrepreneurship and SMMEs.